

**Enquête sur le recrutement
et l'emploi au Québec
(année 2004)**

Rapport méthodologique

Direction de la méthodologie, de la
démographie et des enquêtes spéciales

Institut de la statistique du Québec

Juillet 2005

La méthodologie de l'enquête 3

1 Introduction	4
2 La population visée	4
3 La base de sondage	4
4 L'échantillon	5
5 Le questionnaire	10
6 La collecte des données	11
7 L'estimation.....	16
8 La qualité des estimations	19
9 La confidentialité des résultats.....	24
10 Appréciation globale	25

Annexes..... 27

A) La méthode pour la rotation de l'échantillon.....	28
B) Les indicateurs	30
C) La méthode pour le calcul des proportions d'établissements	33

Liste des tableaux

Tableau 1 Répartition de la population et de l'échantillon d'établissements par secteur d'activité et par taille.....	7
Tableau 2 Renouvellement de l'échantillon pour l'enquête annuelle de 2004	8
Tableau 3 Résultats de la collecte de données par strate	14

LA MÉTHODOLOGIE DE L'ENQUÊTE

1 Introduction

Le Centre d'étude sur l'emploi et la technologie (CETECH) est un centre d'étude dédié au développement d'une expertise en matière de connaissance et d'information sur le marché du travail. Dans l'accomplissement de son mandat, l'enquête sur le recrutement et l'emploi au Québec (EREQ) est un outil d'information. Cette enquête est la première enquête récurrente au Québec et au Canada permettant la production d'indicateurs pour mesurer et anticiper l'évolution du marché du travail dans les secteurs de haute technologie et dans les professions hautement qualifiées. Elle renseigne notamment sur les postes vacants, les départs d'employés, les difficultés de recrutement et les anticipations des employeurs selon le niveau et la nature des qualifications des travailleurs. Cette enquête est réalisée sur une base annuelle. Ce rapport, qui traite de la méthodologie de l'EREQ, portera sur l'année 2004.

À la lecture de ce rapport, on pourra d'abord trouver la description de la population visée, de la base de sondage et de la méthode d'échantillonnage utilisée. Les travaux sur le questionnaire ainsi que les résultats de la collecte des données font l'objet de sections subséquentes. Finalement, les dernières sections de ce rapport présentent une analyse de la qualité statistique des résultats, des risques d'erreur et de la validité des résultats.

2 La population visée

La population visée par l'enquête est constituée de l'ensemble des établissements québécois des secteurs SCIAN de la construction, de fabrication et des services du secteur privé ayant 5 employés ou plus. Les établissements des entreprises publiques provinciales qui sont classifiés dans les trois secteurs mentionnés appartiennent également à la population visée.

3 La base de sondage

La base de sondage utilisée pour l'enquête est la Base de données du registre des entreprises (BDRE) de Statistique Canada mise à jour en juin 2004. Il y a eu un changement de base de sondage par rapport à l'enquête portant sur l'année 2003 puisque c'était la version de décembre 2002 de la BDRE qui avait alors été utilisée. On remarque qu'il y a un an et demi d'écart entre ces deux bases. Ce laps de temps est supérieur à celui entre les bases des enquêtes portant sur les années 2002 et 2003 qui n'était que de 6 mois. Les impacts de ce changement de base seront discutés à la section 4.2. Par rapport à d'autres sources d'informations sur les entreprises, publiques ou privées, la BDRE présente de nombreux avantages. Les plus importants sont sans doute sa mise à jour continue et le fait qu'elle soit constituée spécifiquement pour la réalisation d'enquêtes. Les autres principaux avantages sont la désagrégation de la structure des sociétés jusqu'au niveau de l'établissement et la présence d'une classification industrielle à jour et assez précise. On y trouve aussi une variable catégorielle estimant le nombre d'employés (1 à 4, 5 à 9, etc.) de chacun des établissements, laquelle peut s'avérer précieuse pour l'élaboration du plan de sondage. Il faut aussi noter que depuis la version de juin 2002, la BDRE contient un numéro de téléphone. Ceci représente une amélioration considérable par rapport aux plus anciennes versions. Toutefois, ce numéro correspondant au numéro de l'entreprise et non de l'établissement, il a tout de même fallu procéder à

une recherche de numéros de téléphones pour tous les nouveaux établissements de l'échantillon. Pour les établissements pour lesquels aucun numéro de téléphone n'a été trouvé, l'Institut de la statistique du Québec (ISQ) leur a fait parvenir un formulaire présentant l'enquête et leur demandant de le remplir et de le retourner par télécopieur ou de communiquer avec l'ISQ. Au total, 197 lettres ont été expédiées et 98 d'entre elles, soit 49,8 %, ont rappelé ou ont retourné le formulaire nous fournissant un numéro de téléphone permettant de les rejoindre. En raison des résultats obtenus, cette stratégie est à retenir l'an prochain pour bonifier la base de sondage et diminuer légèrement la non-réponse.

Depuis la version de décembre 2003 de la BDRE, Statistique Canada transmet à l'ISQ l'information sur les emplacements. Les emplacements peuvent être des points de vente ou de service, des entrepôts, etc. relevant d'un établissement. Lors des enquêtes précédentes, suite à une demande spéciale de l'ISQ, une telle liste avait également été transmise par Statistique Canada pour les établissements faisant partie de l'échantillon de l'EREQ seulement. Toutefois, le fait de devoir faire une telle demande occasionnait des délais et limitait la quantité d'information que l'on pouvait recevoir. L'inclusion du niveau emplacement dans la base de sondage constitue donc une amélioration à l'enquête notamment parce qu'il apporte de l'information importante servant à la validation. De plus, pour les nouveaux établissements, on a utilisé les informations de la BDRE comme liste de base d'emplacements à valider avec les répondants. Cette validation a grandement facilité le travail pour s'assurer que le répondant remplissait le questionnaire pour les unités que l'on voulait enquêter.

Malgré ces avantages, il ne faudrait pas croire que la BDRE constitue l'outil parfait pour la réalisation d'enquêtes auprès d'entreprises ou d'établissements. Tout d'abord, la liste des emplacements pour un établissement ne faisant pas l'objet de mises à jour constantes de Statistique Canada, une validation complète de cette liste doit être faite avec les répondants. Cette année toutefois, pour les établissements enquêtés en 2003 et 2004, la validation faite en 2003 a été conservée pour 2004 évitant ainsi aux intervieweurs et aux répondants de reprendre ce travail. D'autre part, on pose l'hypothèse que le fichier contient tous les établissements en activité au moment de sa publication. Cependant, il est inévitable que le fichier contienne également une certaine proportion d'établissements qui ont cessé leurs activités. Par conséquent, il est difficile d'estimer précisément l'ampleur réelle de cette surcouverture. Ainsi, la taille exacte de la population étudiée par l'enquête est inconnue. On reviendra sur cet aspect dans la section traitant de la qualité des estimations.

4 L'échantillon

4.1 La stratification

La population a été stratifiée selon le secteur d'activité économique, défini par le système de classification des industries de l'Amérique du Nord (SCIAN), et selon la taille de l'établissement. Sept secteurs d'activité économique ont été retenus pour la stratification. Notons qu'à l'instar des enquêtes 2002 et 2003, c'est le SCIAN 2002 qui a été utilisé dans cette enquête. Les secteurs retenus se définissent comme suit :

- Services professionnels, scientifiques et techniques (Code du SCIAN : 54)

- Industries manufacturières de haute technologie : produits informatiques et électroniques (Code du SCIAN : 334)
- Industries manufacturières de haute technologie : produits pharmaceutiques et médicaments (Code du SCIAN : 3254)
- Industries manufacturières de haute technologie : produits aérospatiaux et leurs pièces (Code du SCIAN : 3364)
- Autres secteurs de fabrication (Codes du SCIAN : 31, 321-324, 325 excepté 3254, 326, 327, 331-333, 335, 336 excepté 3364, 337, 339)
- Autres secteurs de services (Codes du SCIAN : 41, 44-45, 48, 49, 51-53, 55, 56, 71, 72, 81)
- Construction (Codes du SCIAN 23).

Notons que les établissements des codes du SCIAN 813110, 813310, 813410 et 813910 ont été considérés inadmissibles. Ces codes du SCIAN concernent les organismes religieux, les organismes d'action sociale, les organisations civiques et amicales et les organisations de gens d'affaires. Ces cas pourront être retirés de la population lors du prochain tirage d'échantillon afin d'éviter de les sélectionner et de les déclarer inadmissibles après coup.

À l'intérieur des services professionnels, des autres secteurs de fabrication, des autres secteurs de services et de la construction, on a défini deux strates de taille : l'une contenant les grands établissements et l'autre les petits. La taille est évaluée en fonction du nombre d'employés. Afin de déterminer si un établissement faisait partie de l'une ou l'autre des strates de taille, on procédait de la façon suivante :

- 1 On calculait le nombre total d'employés dans une strate (en additionnant le nombre d'employés de chacun des établissements de cette strate).
- 2 Les n plus grands établissements représentant au moins 50 % des employés de la strate étaient considérés comme faisant partie de la strate des grands établissements. Le reste des établissements était classé dans la strate des petits établissements.

Au total, on dispose donc de onze strates. Cette stratification a pour but de permettre une bonne répartition de l'échantillon par secteur et par taille ainsi que la production d'estimations fiables pour certains secteurs d'activité. Notons qu'il n'y a qu'une seule strate par secteur d'activité pour les industries manufacturières de haute technologie. Puisque ces strates sont recensées, il est inutile de séparer les grands et les petits établissements.

4.2 *La taille et la répartition de l'échantillon*

Lors de l'élaboration de la version annuelle de l'enquête avec le CETECH, une taille d'échantillon de 6 250 établissements était visée. Toutefois, en raison de la méthode retenue pour procéder à la rotation de l'échantillon, le nombre d'établissements enquêtés peut varier légèrement d'une année à l'autre. L'enquête actuelle a été réalisée auprès de 6 334 employeurs québécois. Par ailleurs, les analyses des données dont il sera question dans ce rapport porteront sur 6 646 établissements. Ce nombre comprend les 6 334 établissements enquêtés cette fois-ci ainsi que 312 établissements déclarés inadmissibles de façon permanente lors des réalisations précédentes de l'enquête. Ces établissements étaient alors fermés définitivement ou en double sur la base de sondage. Ces 312

établissements ont dû être conservés pour l'analyse pour des raisons techniques puisqu'ils sont toujours présents sur la base de sondage (BDRE).

La taille et la répartition de l'échantillon d'établissements retenus pour l'année 2004 de l'enquête sur le recrutement et l'emploi au Québec sont présentées au **tableau 1**. Dans ce tableau, on retrouve également la répartition de la population ainsi que la fraction de sondage pour chacune des strates. De plus, dans la dernière colonne du tableau, on donne la proportion d'employés représentée par les grands et les petits établissements de la population dans un secteur d'activité particulier.

Tableau 1
Répartition de la population et de l'échantillon d'établissements par secteur d'activité et par taille

Secteurs d'activité	Taille	Population Établissements (N)	Échantillon Établissements (n)	Fraction de sondage (%)	Proportion d'employés (%)
Services professionnels, scientifiques et techniques	Grands	585	585	100,0	60,3
	Petits	4 623	1 088	23,5	39,7
Fabrication de produits informatiques et électroniques	Grands et petits	295	295	100,0	100,0
Fabrication de produits pharmaceutiques et de médicaments	Grands et petits	65	65	100,0	100,0
Fabrication de produits aérospatiaux et de leurs pièces	Grands et petits	69	69	100,0	100,0
Autres secteurs de fabrication	Grands	1 222	382	31,3	62,1
	Petits	7 579	958	12,6	37,9
Autres secteurs de services	Grands	5 419	679	12,5	53,2
	Petits	50 706	1 077	2,1	46,8
Construction	Grands	1 579	735	46,5	68,5
	Petits	5 102	713	14,0	31,5
Total		77 244	6 646	8,6	

Il avait été prévu, au moment de la planification de la version annuelle de l'enquête, qu'une rotation de l'échantillon aurait lieu entre l'enquête portant sur l'année 2003 et celle portant sur l'année 2004. Cette rotation a pour objectif principal de permettre, de façon contrôlée, aux établissements de sortir de l'enquête après un certain temps, autrement qu'en fermant leurs portes. Naturellement, cette rotation ne peut s'appliquer qu'aux sept strates échantillonnées de la construction et des secteurs de fabrication et de services, les strates recensées ne pouvant faire l'objet d'une rotation. Le fait que les nouveaux établissements puissent être échantillonnés en plus grand nombre est un autre atout de la rotation de l'échantillon puisque ce dernier est alors plus représentatif de la population actuelle.

Plusieurs méthodes existent pour faire la rotation d'un échantillon. Celle qui a été utilisée dans cette enquête est appelée échantillonnage colloqué modifié¹ (en anglais : « modified collocated sampling »). Cette méthode a donc été appliquée pour tirer l'échantillon dans chacune des sept strates échantillonnées. Une description détaillée et un exemple de cette méthode se trouvent à l'annexe A.

Tableau 2
Renouvellement de l'échantillon pour l'enquête annuelle de 2004¹

Secteurs d'activité	Taille	Échantillonnés pour l'enquête actuelle (n)	Échantillonnés pour l'enquête précédente		Non échantillonnés pour l'enquête précédente (n)
			Admissibles et inadmissibles temporaires (n)	Inadmissibles permanents ² (n)	
Services professionnels, scientifiques et techniques	Grands	585	361	80	144
	Petits	1 088	517	35	536
Fabrication de produits informatiques et électroniques	Grands et petits	295	216	25	54
Fabrication de produits pharmaceutiques et de médicaments	Grands et petits	65	46	9	10
Fabrication de produits aérospatiaux et de leurs pièces	Grands et petits	69	52	7	10
Autres secteurs de fabrication	Grands	382	250	17	115
	Petits	958	566	30	362
Autres secteurs de services	Grands	679	385	38	256
	Petits	1 077	640	36	401
Construction	Grands	735	341	24	370
	Petits	713	353	11	349
Total		6 646	3 727	312	2 607

¹ L'enquête actuelle est l'enquête annuelle de 2004 et l'enquête précédente est l'enquête annuelle de 2003.

² Les inadmissibles permanents ont été échantillonnés lors des enquêtes semestrielles de 2001 et 2002 ainsi que lors des enquêtes annuelles de 2002 et 2003 et ont alors été déclarés inadmissibles permanents.

Le **tableau 2** présente le renouvellement de l'échantillon pour l'enquête de 2004. En effet, on y retrouve la répartition des établissements échantillonnés pour l'enquête de 2004 selon leur statut à l'enquête de 2003. On s'aperçoit, en regardant ce tableau, que 2 607 établissements n'ayant pas été échantillonnés lors de l'enquête annuelle 2003 l'ont été lors de l'enquête annuelle 2004. Ces nouveaux établissements sont entrés dans l'échantillon en raison de la rotation ou pour remplacer les établissements morts. Par établissement mort, on entend un établissement qui était présent sur l'ancienne base de sondage, mais qui ne l'est plus sur la nouvelle parce qu'il a fermé ses portes

¹ K.P. Srinath et Ronald M. Carpenter, *Buisness Survey Methods*, Wiley & sons, 1995, chapitre 10, pp. 171-183

entre temps, un établissement fusionné ou encore un établissement qui a changé de strate. Dans ce dernier cas, l'établissement est un mort dans une strate et un nouveau dans l'autre. Pour cette enquête, il avait été établi que les établissements morts seraient remplacés de même que 20% des établissements répondants pour l'année 2003, ceci permet d'assurer la rotation.

Le nombre de nouveaux établissements est plus élevé cette année que ce que l'on avait prévu au départ. En fait, dans l'échantillon de 2004, il y a plus d'établissements non échantillonnés en 2003 qu'il y en avait de non échantillonnés en 2002 dans l'échantillon de l'enquête 2003, alors qu'il y avait pourtant eu l'ajout du secteur de la construction. Plusieurs raisons expliquent ce phénomène. Tout d'abord, l'écart d'un an et demi entre les versions de la BDRE utilisées en 2003 et 2004 étant plus grand qu'à l'habitude, les morts et les naissances relevées sur la nouvelle base sont plus nombreuses que lors des enquêtes précédentes. Deuxièmement, la taille de la population a diminué dans presque toutes les strates. Ceci entraîne une augmentation de la fraction de sondage et donc l'échantillonnage de plus de nouveaux établissements. Enfin, comme le nombre d'établissements à échantillonner dans la strate des petits établissements des services professionnels, scientifiques et techniques dépend du nombre d'établissements dans les strates recensées et que ce dernier a diminué cette année, la fraction de sondage de cette strate a augmenté amenant encore une fois plusieurs nouveaux établissements.

Ensuite, à la cinquième colonne du **tableau 2**, on constate que 312 établissements échantillonnés pour l'année 2004 avaient déjà été enquêtés lors d'une des enquêtes précédentes et ils avaient alors été déclarés inadmissibles de façon permanente. Ces 312 établissements n'ont pas été rappelés par les interviewers lors de la présente enquête, mais ils ont été conservés dans les analyses parce qu'ils sont toujours présents sur la base de sondage (c'est une exigence de l'échantillonnage colloqué modifié). Enfin, la quatrième colonne du **tableau 2** montre qu'il y a 3 727 établissements admissibles ou temporairement inadmissibles (établissements possédant moins de 5 employés ou étant temporairement fermés) lors de l'enquête 2003 qui sont à nouveau dans l'échantillon de l'enquête 2004.

De la population de 77 244 établissements, 6 646 ont donc été sélectionnés pour faire partie de l'échantillon, selon la structure illustrée au **tableau 1**. Tout comme lors des enquêtes précédentes, pour les trois secteurs de fabrication de haute technologie, les établissements étaient recensés. Il en était de même pour les grands établissements du secteur des services de haute technologie. À l'intérieur des sept autres strates, l'échantillonnage colloqué modifié a permis de sélectionner l'échantillon. Il est à noter que les grands établissements des autres secteurs de fabrication, ceux des autres secteurs de services et ceux du secteur de la construction n'ont pas été recensés. Cela s'explique par la grande taille de la population dans ces strates (respectivement 1 222, 5 419 et 1 579 établissements).

4.3 La répartition régionale de l'échantillon

Depuis l'enquête annuelle de 2002, des estimations sont produites pour trois regroupements de régions, soient:

- Région de Montréal : Montréal, Laval, Lanaudière, Laurentides et Montérégie
- Régions ressources : Bas-Saint-Laurent, Saguenay-Lac-Saint-Jean, Mauricie, Abitibi-Témiscamingue, Côte-Nord, Nord-du-Québec, Gaspésie-Iles- de- la –Madeleine

- Régions centrales : Estrie, Centre-du-Québec, Chaudière-Appalaches, Capitale-Nationale, Outaouais.

Ainsi, un établissement possédant des emplacements dans deux ou trois de ces régions devait compléter l'ensemble du questionnaire deux ou trois fois selon le cas. Pour saisir cette information, des enregistrements supplémentaires ont été ajoutés au fichier de l'échantillon qui contenait à l'origine 6 646 établissements. En fait, 182 établissements (parmi les 6 334 transmis à la Direction des activités de collecte de l'ISQ) étaient formés de plus d'un emplacement et étaient donc potentiellement présents dans plusieurs régions.

Ainsi, une fois la collecte terminée et la liste d'emplacements mise à jour par les répondants, le fichier final contient 6 732 observations : 6 152 établissements ayant un seul emplacement, 90 ayant plusieurs emplacements tous concentrés dans la même région, 42 ayant des emplacements dans deux régions, 22 ayant des emplacements dans trois régions, 28 ayant plusieurs emplacements mais étant non-répondants ou inadmissibles (dans ce cas, une seule observation apparaît au fichier) et 312 inadmissibles permanents.

5 Le questionnaire

Le questionnaire de la version 2004 de l'enquête est presque identique à celui de la version 2003. Toutefois, suite à la collecte de 2003, le Centre d'étude sur l'emploi et la technologie (CETECH) et l'ISQ ont proposé quelques légères modifications visant à faciliter le traitement des données et ont collaboré à la production de la version finale du questionnaire. Les questions 1 et 10 ont notamment été légèrement modifiées. Les résultats de ces modifications sur le terrain seront discutés à la section 6.1. Cette nouvelle version est composée de 11 questions plutôt que 10 comme c'était le cas en 2003. Ainsi, une nouvelle question portant sur le recours à l'apprentissage en ligne comme moyen de formation a été ajoutée. L'ensemble des thèmes qui sont abordés dans le questionnaire de l'enquête annuelle 2004 sont les suivants:

1. Nombre de personnes qui travaillaient dans l'établissement au 31 décembre 2004
2. Embauche de nouveaux employés entre le 1^{er} janvier et le 31 décembre 2004
3. Difficultés de recrutement pour l'année 2004
4. Départs d'employés entre le 1^{er} janvier et le 31 décembre 2004
5. Postes actuellement vacants
6. Postes vacants depuis au moins quatre mois
7. Mesures prises pour faire face à la difficulté que posent ces postes demeurés longtemps vacants
8. Embauche de personnes provenant de l'extérieur du Québec entre le 1^{er} janvier et le 31 décembre 2004
9. Perspectives d'affaires de l'établissement pour l'année 2005
10. Variation du nombre d'employés à prévoir pour l'année 2005
11. Recours à l'apprentissage en ligne comme moyen pour dispenser de la formation.

Les répondants doivent fournir de l'information sur chacun de ces thèmes non seulement pour l'ensemble de leurs employés, mais également pour chacune des catégories d'emploi présentes dans leur établissement (sauf pour les questions 7, 9 et 11 qui ne concernent que l'établissement).

Ces catégories et les professions associées à chacune d'elle sont tirées de la classification nationale des professions. Les catégories d'emploi sont les suivantes :

1. Cadres
2. Professionnels des sciences naturelles et appliquées
3. Autres professionnels
4. Techniciens des sciences naturelles et appliquées
5. Autres techniciens
6. Secrétaires
7. Métiers
8. Superviseurs et surveillants (vente ou production)
9. Personnel général de bureau
10. Personnel général de la vente et des services
11. Travailleurs de la production ou domaine connexe (transport, installation, entretien).

Depuis l'année dernière, un questionnaire particulier a été fait pour les établissements ayant plusieurs emplacements. L'établissement reçoit le nombre de questionnaires correspondant au nombre de régions dans lesquelles se trouvent ses emplacements. La première page de chaque questionnaire indique la région en question. De plus, la liste des emplacements correspondants à chaque région apparaît à la deuxième page du questionnaire. Ceci facilite grandement la vie des répondants. De plus, les répondants mettent à jour la liste de leurs emplacements, ce qui nous aidera également pour les prochaines enquêtes.

Il est à noter que le questionnaire cherche à exclure l'effet de saisonnalité. En effet, les questions 2, et 4 du questionnaire mentionnent aux répondants qu'ils ne doivent pas tenir compte des employés mis à pied puis réengagés durant la même année.

6 La collecte des données

6.1 Le déroulement

À l'instar de ce qui avait été fait pour l'enquête 2003, un filtre téléphonique a été fait auprès de tous les nouveaux établissements et auprès des établissements provenant d'entreprises dont la structure est complexe. Ce filtre permet de vérifier l'admissibilité de l'établissement et de s'assurer de l'adresse d'expédition. Il s'est déroulé entre le 25 janvier et le 15 février. Une innovation a été amenée cette année, à savoir le filtre des nouveaux établissements ayant plusieurs emplacements. Pour ces derniers, en plus des vérifications de l'admissibilité et des coordonnées, une vérification de la liste des emplacements a été faite afin d'imprimer la liste à jour sur le questionnaire. Ainsi, la validation des emplacements n'a pas eu à se faire pendant la collecte comme cela a été le cas la dernière fois. Dans le cas des établissements enquêtés en 2003 toutefois, la liste d'emplacements mise à jour l'an passé était revue seulement au moment de la collecte.

La collecte des données de l'enquête s'est ensuite déroulée du 8 mars au 20 mai 2005. Les heures de collecte se sont étendues entre 8 h 30 et 16 h 30, du lundi au vendredi. Des périodes de collecte ont également été prévues de 12h à 13h pour les répondants difficiles à rejoindre. Il y a également

eu trois périodes de collecte de 16h30 à 20h30 pour les dossiers sur rendez-vous ainsi que pour les dossiers avec qui il n'y avait pas eu de contacts durant le jour.

Les entrevues ont été faites à l'aide d'un logiciel de type ITAO (interview téléphonique assistée par ordinateur) par une équipe de 14 interviewers, sous la supervision constante d'une chargée d'enquête. Les interviewers avaient reçu au préalable une formation de deux jours portant sur les objectifs de l'enquête, le contenu du questionnaire, la procédure de validation des coordonnées des entités enquêtées et les codes de résultats. La formation comportait une partie théorique, des mises en situation, la présentation du questionnaire informatisé et une période de mise en pratique pour se familiariser avec les outils de collecte, les catégories professionnelles ainsi que la documentation disponible.

Tous les établissements ont fait l'objet d'une validation des coordonnées (nom et adresse) afin de s'assurer que l'unité échantillonnée était bien celle enquêtée. La procédure de validation utilisée a été développée conjointement par deux directions de l'ISQ : celle des services et stratégies de collecte et celle de la méthodologie, de la démographie et des enquêtes spéciales. Elle permet de colliger les changements de données et de déterminer si l'établissement demeure admissible à l'enquête. Cette procédure a permis d'uniformiser les validations faites par les interviewers.

En moyenne, six appels à des heures et à des jours différents ont été effectués pour rejoindre la personne sélectionnée au sein de l'établissement et pour compléter l'entrevue. Ce nombre diminue à 4 pour les inadmissibles et monte à 11 pour les dossiers demeurés sans réponse. Pour les établissements ayant plusieurs emplacements, sept appels ont été faits en moyenne. Notons que le nombre d'appels a varié entre 1 et 32. La réalisation des entrevues s'est effectuée en français ou en anglais, au choix du répondant.

Soulignons qu'afin d'aider les répondants à répartir leurs employés entre les divers groupes de professions, une annexe détaillée a été utilisée par les interviewers. Ce document était beaucoup plus détaillé que celui remis aux répondants et il leur permettait d'intervenir de façon efficace auprès des répondants et d'obtenir une classification plus sûre. À cet égard, le fichier remis par le CETECH contenant la liste des professions et leur répartition à l'intérieur de la classification nationale des professions a été fort utile.

Lors de l'enquête 2003, certains problèmes de compréhension des répondants étaient survenus à la question 1 en raison des deux périodes de référence auxquelles elle se rapporte à savoir le nombre de personnes en emploi au 31 décembre et l'existence des catégories professionnelles au cours de l'année de l'enquête. Pour éliminer cette ambiguïté, la présentation du tableau a été modifiée et des explications additionnelles ont été inscrites. Ces modifications ont eu un effet bénéfique sur la compréhension des répondants et devraient être maintenues. L'ajout d'un exemple pourrait aider encore davantage les répondants, notamment dans le secteur de la construction. Par ailleurs, l'ajout à la question 1 de la précision sur les employés mis à pied après le 15 décembre et réembauchés avant le 15 janvier a permis d'éviter les problèmes que cette pratique avait causé en début de collecte l'année dernière et de s'assurer que tous les répondants incluaient bien ces employés lorsqu'ils remplissaient leur questionnaire.

Bien que le questionnaire soit en général bien compris par les répondants, certaines améliorations pourraient encore y être apportées. Ainsi, certains concepts pourraient être définis davantage pour

faciliter la vie des répondants et des intervieweurs. C'est notamment le cas des étudiants et des employés mis à pied puis réembauchés durant la même année. La notion de poste vacant a aussi causé quelques difficultés puisque les répondants ne lisent pas toujours la note expliquant ce que l'on entend par poste vacant. Des erreurs se glissent donc à cette question et ne sont détectées qu'aux questions subséquentes, ce qui oblige les intervieweurs à revenir en arrière. Une façon de régler ce problème pourrait être de diviser la question de manière à demander d'abord s'il y a des postes vacants puis à demander si ces postes ont donné lieu à des activités de recrutement et sont ouverts à des personnes de l'extérieur de l'établissement.

Depuis la précédente édition de l'enquête, la présentation de la question 4 a été modifiée afin de placer le nombre total de départs au début plutôt qu'à la fin du tableau. Tout comme c'était le cas pour l'enquête de 2003, ce déplacement a occasionné plusieurs erreurs de la part des répondants malgré le fait que la ligne du nombre total de départs ait été tramée cette année afin de la rendre plus visible. Il faudrait réévaluer la pertinence de cette disposition pour la prochaine enquête. Il faut également souligner que la question 4b a semblé difficile à répondre par les répondants parce que l'information ainsi désagrégée n'est pas souvent informatisée. La qualité des données peut être moins fiable pour cette question. Enfin, la question 10 a été modifiée cette année. En effet, la réponse « n'existera pas » a été remplacée par « n'existe pas » pour éviter d'avoir à recoder des réponses lors de l'estimation. Ce changement a permis de faciliter le travail d'estimation dont il sera davantage question à la section 7.

Par ailleurs, tout au long de la collecte, des contrôles de qualité ont été effectués afin de s'assurer du respect des normes de confidentialité et de qualité de l'Institut. À titre d'exemple, une validation historique a été faite pour comparer le nombre d'employés donné par le répondant à celui donné l'année dernière s'il était répondant. Cette validation a été faite pour tous les établissements, à l'exception de ceux qui avaient fourni des données consolidées lors de la précédente enquête. De façon générale, mis à part les quelques points mentionnés précédemment, aucun problème particulier n'est venu perturber le déroulement de la collecte téléphonique.

6.2 Les résultats

Au total, les entrevues téléphoniques ont été complétées pour 4 544 établissements ce qui donne un taux de réponse de 81,7 % (80,6 % lors de l'enquête 2003). Cette hausse du taux de réponse est intéressante puisqu'elle survient malgré l'introduction d'un peu plus de nouveaux établissements que l'an dernier. Notons que le taux de réponse, qui avait subi une légère baisse lors des deux premières enquêtes annuelles, est maintenant revenu au niveau de celui obtenu dans les enquêtes semestrielles (81,7% lors du semestre de janvier à juin 2002). Le nombre d'enregistrements servant dans les calculs d'estimation est 4 630 (4 480 enregistrements provenant d'établissements présents dans une seule région, 84 provenant d'établissements présents dans deux régions et 66 provenant d'établissements présents dans trois régions). Pour les établissements ayant un seul emplacement, la durée moyenne pour compléter une entrevue était de 7 minutes (elle était de 6 minutes en 2003), la durée maximale de 40 minutes et la durée minimale de 2 minutes. La durée maximale se produisait lors de situations plus complexes, comme par exemple pour les entreprises possédant plusieurs employés ou pour des fusions récentes. Par ailleurs, pour les établissements ayant plusieurs emplacements, la durée moyenne pour compléter une entrevue était de 22 minutes (elle était de 11 minutes en 2003), avec un temps maximum de 94 minutes et un temps minimum de 2 minutes. L'augmentation du temps d'entrevue pour ces établissements est principalement causée

par l'intégration au questionnaire informatisé de la validation des emplacements. Ce travail se faisait de façon papier et manuelle en 2003 et le temps de cette tâche n'était donc pas comptabilisé dans l'entrevue.

La collecte s'est soldée par un taux de collaboration de 98,3 %. L'année dernière, ce taux se situait à 97,9 %. Le taux de collaboration est défini par :

$$\text{Taux de collaboration} = \frac{\text{entrevues complétées}}{\text{entrevues complétées} + \text{refus}} .$$

Dans la formule ci-dessus, les refus peuvent provenir d'un manque de disponibilité, d'un refus de donner accès à la personne sélectionnée après plusieurs tentatives ou encore d'un refus catégorique. Soulignons que les établissements les plus importants pour la qualité des données ont bien collaboré à l'enquête et ont répondu massivement. L'enquête a permis de repérer 772 établissements (parmi les 6 334 contactés par les interviewers de l'ISQ) qui n'étaient pas admissibles à l'enquête : 503 où il y a moins de 5 employés, 189 fermés définitivement, 9 fermés temporairement, 3 établissements fusionnés, 55 hors population et 13 en double. Toutefois, pour compléter l'échantillon d'analyse (composé de 6 646 établissements), on doit aussi considérer les 312 inadmissibles détectés lors des enquêtes précédentes. Le nombre total d'établissements inadmissibles pour l'enquête actuelle est donc de 1 084.

Le **tableau 3** présente les résultats de la collecte pour chacune des strates de l'enquête. Dans ce tableau, on constate que les taux de réponse par strate sont tous supérieurs à 75 % et qu'ils sont même supérieurs à 80 % dans huit strates (ce n'était le cas que dans cinq strates lors de l'enquête précédente). On constate que les hausses des taux de réponse sont particulièrement appréciables dans la strate de fabrication de produits pharmaceutiques et de médicaments ainsi que dans la strate de fabrication de produits aérospatiaux et de leurs pièces (respectivement 8,9 % et 6,6 % d'augmentation par rapport à 2003). Des efforts considérables sont mis pour obtenir de bons taux de réponse dans ces strates névralgiques pour l'enquête et il semble que cela porte fruit. On note également que la hausse du taux de réponse global est perceptible dans presque toutes les strates. En effet, tous les taux de réponse par strate ont augmenté par rapport à l'enquête sur l'année 2003, sauf dans les strates des autres secteurs de services et dans la strate des petits établissements de la construction où le taux de réponse a diminué très légèrement. Les taux de réponse par strate varient entre 77,5 % et 88,9 %.

Tableau 3
Résultats de la collecte de données par strate

Secteurs d'activité	Taille	Échantillon (n)	Échantillon admissible (n)	Répondants (n)	Taux de réponse (%)
Services professionnels, scientifiques et techniques	Grands	585	473	383	81,0
	Petits	1 088	814	659	81,0
Fabrication de produits informatiques et électroniques	Grands et petits	295	247	207	83,8
Fabrication de produits pharmaceutiques et de médicaments	Grands et petits	65	54	48	88,9
Fabrication de produits aérospatiaux et de leurs pièces	Grands et petits	69	57	48	84,2
Autres secteurs de fabrication	Grands	382	359	285	79,4
	Petits	958	861	720	83,6
Autres secteurs de services	Grands	679	573	444	77,5
	Petits	1 077	840	660	78,6
Construction	Grands	735	653	562	86,1
	Petits	713	631	528	83,7
Total		6 646	5 562	4 544	81,7

En ce qui a trait à la non-réponse partielle, seulement un petit nombre de cas ont été rencontrés. En effet, pour 107 établissements (2,4 % des répondants), des réponses manquaient à certaines questions. Voici les différents types d'imputation utilisées. Dans 76 cas, il a fallu procéder à une imputation à l'aide des informations de la BDRE (informations sur les entreprises et sur les établissements) pour répartir le nombre d'employés d'une entreprise entre ses divers établissements. Dans certains de ces cas, le répondant fournissait des informations pour des établissements qui n'avaient pas été enquêtés en surplus. De plus, 7 établissements présents dans plusieurs régions ont donné un chiffre global pour l'ensemble du Québec et il a fallu le répartir dans chacune des régions. Ici encore, l'information de la BDRE a été utilisée. D'autres répondants ont donné des réponses distinctes pour chaque emplacement même s'il y en avait plusieurs dans la même région. Une imputation a alors été nécessaire dans 4 cas pour regrouper ces informations. Pour 8 établissements, une imputation déterministe a été nécessaire pour corriger les données suite à la validation. Enfin, une nouvelle forme de non-réponse est apparue cette année pour les questions 4 et 5. Ainsi, 12 établissements ont fourni seulement des totaux sans les répartir dans les catégories professionnelles ou dans les raisons de départ. Une imputation déterministe ou par donneurs a alors été nécessaire pour régler ces cas. Notons que le nombre de cas de non-réponse partielle a augmenté par rapport aux deux enquêtes précédentes alors qu'il y en avait eu respectivement 78 et 66 cas.

On remarque que pour le secteur de fabrication de produits aérospatiaux et de leurs pièces, il y a 11 cas d'imputation sur les 48 établissements répondants (22,9 % des répondants de ce secteur). Il est à noter que presque autant d'imputation avait également été faite pour cette strate lors de l'enquête précédente (10 cas sur 45 répondants). Par ailleurs, une strate qui a connu une forte augmentation du nombre de cas d'imputation est celle des grands établissements des services professionnels, scientifiques et techniques. Ainsi, l'an dernier, une imputation a été nécessaire dans 4,6 % des répondants de cette strate alors que cette proportion a grimpée à 9,1 % cette année (35 cas sur 383 répondants). Dans les autres strates, la proportion de cas imputés est toujours inférieure à 5 %.

Il y a lieu de se questionner sur les impacts potentiels d'une augmentation des cas d'imputation. Une brève analyse a donc été faite consistant à comparer les répondants complets et partiels selon différentes caractéristiques. Pour ce faire, des tests du khi-deux ont été faits afin de détecter l'association entre le statut de réponse (complète ou partielle) et la caractéristique à l'étude. Il en est ressorti que l'imputation est principalement concentrée dans les deux strates mentionnées précédemment plutôt que dans les autres strates, dans les établissements de grande taille plutôt que dans les petits, dans les établissements ayant plusieurs emplacements plutôt que dans les établissements uniques, dans les établissements faisant partie d'entreprises complexes plutôt que dans les établissements formant une entreprise en soi ainsi que dans les établissements des régions Montréal et centrales plutôt que dans les établissements des régions ressources. Plusieurs de ces variables sont vraisemblablement liées, comme la complexité de l'entreprise à laquelle appartient l'établissement et la région. Ainsi, il y a probablement moins d'entreprises complexes dans les régions ressources que dans les régions centrales et de Montréal. Les analyses présentées ici dressent donc un portrait sommaire de la situation. Pour une analyse plus approfondie, il faudrait procéder à des analyses multivariées.

Un point rassurant malgré l'augmentation du nombre de cas de non-réponse partielle est que dans la grande majorité des cas, les répondants fournissent des totaux et il nous faut les répartir (entre les divers établissements d'une entreprise ou entre les diverses régions). Les estimations de totaux ne devraient donc pas être affectées par un biais, mais la variance peut être sous-estimée. Ainsi, lorsqu'il y a un grand nombre de cas d'imputation, on peut obtenir une augmentation de la variance de l'estimateur qui n'est pas mesurée par le système. Toutefois, dans ce cas-ci, comme relativement peu d'imputation a été nécessaire, il n'y a pas lieu de s'inquiéter de ce phénomène, sauf dans les secteurs de fabrication de produits aérospatiaux et de leurs pièces ainsi que les grands établissements des services professionnels, scientifiques et techniques où le nombre de cas d'imputation est plus important. Lors de la prochaine enquête, il y aurait lieu de prendre des mesures afin d'éviter que le nombre de cas d'imputation ne continue d'augmenter et ne devienne problématique. Par exemples, des rencontres avec certains répondants importants pourraient permettre de trouver une façon de simplifier leur travail tout en limitant le plus possible le travail d'imputation, une sensibilisation accrue pourrait être faite auprès de l'ensemble des répondants afin de leur faire prendre conscience de l'importance de données précises et d'éviter que le fait de donner des données consolidées ne devienne la voie de la facilité, etc.

7 L'estimation

7.1 La pondération

La pondération a pour but fondamental de permettre l'inférence des résultats observés à la population étudiée en respectant le plan de sondage qui a permis de recueillir les données. En fait, le poids accordé à un établissement reflète le nombre d'établissements qu'il « représente » dans la population. Ce poids est donc fonction de deux réalités. Initialement, il représente l'inverse de la probabilité de sélection d'un établissement dans sa strate. Ensuite, dans le calcul du poids final, il faut aussi tenir compte de la structure de la réponse observée dans la strate à laquelle l'établissement appartient, **en supposant que le profil des non-répondants de cette strate est semblable au profil des répondants**.

Par ailleurs, pour effectuer une pondération adéquate, il faut être en mesure d'estimer le nombre total d'établissements inadmissibles de la population. Dans le cadre de la présente enquête, on a obtenu 4 544 répondants admissibles, 1 084 répondants inadmissibles (dont 312 identifiés lors d'une enquête précédente) et 1 018 non-répondants. Seuls ces derniers possèdent un statut inconnu. Pour cette raison, l'hypothèse suivante a été utilisée : parmi les 1 018 non-répondants à l'enquête, certains peuvent être considérés comme inadmissibles. On peut déterminer un nombre (i_{nr}) de non-répondants inadmissibles d'une part, en examinant la nature de la non-réponse par strate et d'autre part, en utilisant la répartition des répondants inadmissibles à l'intérieur de chacune des strates. Une fois déterminé, ce nombre plus le nombre de répondants inadmissibles constituent le nombre total d'établissements inadmissibles de l'échantillon.

Soit N_h = nombre total d'établissements de la strate h .

Soit n_h = nombre d'établissements sélectionnés pour faire partie de l'échantillon de la strate h .

Soit r_h = nombre d'établissements répondants de la strate h .

Soit nr_h = nombre d'établissements non-répondants de la strate h .

Soit i_{nr_h} = nombre d'établissements non-répondants considérés comme inadmissibles de la strate h .

Alors le poids Wr_h accordé aux **répondants** de la strate h se calcule ainsi :

$$Wr_h = \underbrace{\frac{N_h}{n_h}}_{\text{Inverse de la prob. de sélection}} \times \underbrace{\frac{r_h + nr_h - i_{nr_h}}{r_h}}_{\text{Ajustement pour la non réponse de la strate h}}$$

7.2 Ajustement de non-réponse particulier

Dans certaines strates, il se peut que selon leurs caractéristiques, les établissements n'aient pas tous répondu en proportion identique. Par exemple, des établissements de la région de Montréal auront peut-être moins répondu que ceux provenant d'une autre région du Québec. Pour cette raison, il est opportun d'examiner la structure de réponse pour chacune des strates. Afin d'effectuer

cette analyse de non-réponse par strate, on sélectionne tout d'abord des variables ou caractéristiques selon lesquelles la non-réponse pourrait varier. Ensuite, on crée, pour chacune des variables choisies, un tableau de fréquences à deux dimensions croisant le statut de réponse (répondant ou non-répondant) d'un établissement avec la variable choisie. Pour tous les tableaux créés, on évalue alors l'association ou la dépendance entre le statut de réponse et la variable choisie. Ceci peut être fait à l'aide de la technique connue sous le nom de « chi-square automatic interaction detection (CHAID) ». Cette technique consiste à effectuer plusieurs tests statistiques du khi-deux afin d'obtenir les différentes associations entre le statut de réponse et les variables choisies. De cette façon, il a été possible de déterminer, dans le cas de la présente enquête, que dans la strate des petits établissements des services professionnels, scientifiques et techniques, dans les strates des établissements des autres secteurs de fabrication et de services ainsi que pour la strate des grands établissements du secteur de la construction, la probabilité de répondre était différente pour les établissements de la région de Montréal comparativement à ceux des autres régions du Québec. Ainsi, les taux de réponses obtenus dans ces six strates pour les établissements de la région de Montréal sont de 82 %, 76 %, 83 %, 75 %, 77 % et 85 % tandis que ceux pour les établissements des autres régions sont respectivement de 89 %, 87 %, 89 %, 88 %, 85 % et 92 %. D'autre part, dans la strate des grands établissements des services professionnels, scientifiques et techniques, le taux de réponse varie selon une combinaison de deux variables. Ainsi, parmi les établissements des secteurs du SCIAN 5413, 5415 et 5417 de la région de Montréal, le taux de réponse est de 75 % alors qu'il est de 89 % parmi les établissements des mêmes secteurs mais dans les autres régions. Enfin, chez les établissements des autres secteurs du SCIAN 541, le taux de réponse est de 87 % toutes régions confondues. En conséquence, l'ajustement de non-réponse dans toutes ces strates diffère selon la provenance de l'établissement et/ou le secteur d'activité économique selon le cas.

7.3 La méthode d'estimation

Le plan de sondage qui a été utilisé dans le cadre de l'enquête est un plan stratifié, où la sélection d'un échantillon s'est faite par échantillonnage colloqué modifié, dans la strate de petits établissements du secteur des services professionnels, scientifiques et techniques ainsi que dans les six strates des autres secteurs de fabrication, de services et de la construction. Pour ce qui est des autres strates, un recensement a été effectué.

Pour cette étude, les estimations suivantes sont produites :

- totaux pour les questions 1, 2b, 4b, 5b, 6b et 8b
- proportions d'établissements pour les questions 2a, 3, 4a, 5a, 6a, 6b, 7, 8a, 9, 10, 11a et 11b
- quotients pour les questions 2a, 3, 4a, 5a, 6a, 8a, 9, 10 et 11a
- indicateurs pour les questions 9 et 10 (voir l'annexe B pour des détails sur les indicateurs).

Toutes ces estimations sont regroupées dans une banque de données. Les marges d'erreur (voir section 8.1) qui accompagnent ces estimations sont calculées, selon le plan de sondage, à l'aide des logiciels SAS et SUDAAN. Soulignons qu'aucun test de comparaison de moyenne n'a été produit. Des tests de Fisher avec l'ajustement de Satterthwaite ont déjà été produits dans cette enquête lorsque les résultats étaient présentés sous forme de tableaux papiers. Toutefois, ils ont été abandonnés depuis la production exclusive de la banque de données en 2003.

Tel que décrit à la section 4.3, le fichier de données utilisé pour produire les estimations contient plus d'observations qu'il y a d'établissements. La façon d'estimer les proportions doit être adaptée. La méthode de calcul des proportions est détaillée à l'annexe C.

8 La qualité des estimations

Tout résultat tiré d'enquêtes ou de recensements est potentiellement entaché d'erreurs. Ces erreurs se répartissent en deux catégories. D'une part, il y a les erreurs dues à l'échantillonnage. Il s'agit des erreurs découlant du fait que seule une partie de la population visée est enquêtée par l'entremise d'un échantillon aléatoire. Ces erreurs sont contrôlables a priori dans le cadre de la définition du plan de sondage et de la détermination de la taille d'échantillon. De plus, elles sont quantifiables a posteriori et permettent de prendre en considération la variabilité des estimations produites lors de leur analyse.

D'autre part, il y a les erreurs qui ne sont pas dues à l'échantillonnage. Parmi celles-ci, se trouvent les différents types de biais pouvant influencer sur les résultats : biais dus à la non-réponse, biais dus aux activités ou au mode de collecte, biais de réponses, surcouverture ou sous-couverture de la population par la base de sondage, etc. Les erreurs non dues à l'échantillonnage ne sont pas formellement quantifiables comme le sont les erreurs de la première catégorie. Il est donc important de tenter le plus possible de détecter leur présence et d'évaluer, le cas échéant, leur ampleur et leur impact.

On peut statuer sur la qualité des estimations en examinant certaines statistiques notamment la précision et la fiabilité des estimations. Les taux de réponse, au total et par question, permettent en outre de porter un jugement éclairé sur la fiabilité des résultats obtenus. Par ailleurs, la marge d'erreur et le coefficient de variation (C. V.) d'une estimation permettent de juger de la précision de celle-ci. En fonction de ces observations, l'utilisateur peut juger de l'importance des erreurs de mesure, des risques de biais et de la portée possible de l'inférence statistique.

8.1 Les erreurs dues à l'échantillonnage

Les erreurs liées à l'échantillonnage sont mesurées à l'aide de la marge d'erreur et du coefficient de variation.

La marge d'erreur (m.e.) est une mesure de précision de l'estimation et se définit comme suit :

$$\text{m.e.} = z_{1-\alpha/2} \sqrt{(\text{variance de l'estimation})}$$

où $z_{1-\alpha/2}$ est une valeur qui est telle que la probabilité qu'une variable aléatoire provenant de la loi normale de moyenne 0 et de variance 1 soit supérieure à $z_{1-\alpha/2}$ est $\alpha/2$. Avec un seuil $\alpha = 0.05$, cette valeur est 1,96. À partir de la marge d'erreur, il est possible de définir l'intervalle de confiance (I.C.) à 95 % qui est associé à l'estimation :

$$\text{I.C.} = \text{estimation} \pm \text{m.e.}$$

Cet intervalle illustre l'étendue des valeurs possibles que peut prendre la variable étudiée dans la population observée. Cela signifie que, si l'on refaisait le sondage un très grand nombre de fois, 19 intervalles sur 20 contiendraient la valeur réelle de la proportion estimée. On doit noter que la marge d'erreur se calcule selon l'hypothèse que les moyennes ou les totaux estimés suivent une loi normale. Ce résultat, qui est asymptotiquement démontré par le théorème de la limite centrale, peut ne pas s'appliquer dans le cas où la taille de l'échantillon est petite et que la distribution des valeurs mesurées est très étendue ou dans le cas de l'estimation d'une petite proportion (moins de 5 %). Tel que mentionné à la section 7.3, des quotients sont produits dans le cadre de cette enquête. Or, l'estimateur d'une telle quantité est biaisé². Pour cette raison, lors du calcul de la marge d'erreur associée à l'estimateur d'un quotient, on tient compte de la variance de l'estimateur et de la valeur maximale que peut atteindre le biais de l'estimateur³.

Le coefficient de variation (C.V.) est une mesure qui permet de quantifier la précision de l'estimation et il se définit comme suit :

$$C.V. = \frac{\text{marge d'erreur}}{(1.96 * \text{estimation})} .$$

Cette mesure contribue à faciliter l'interprétation quant à la précision d'une estimation. Plus le C.V. est élevé, moins précise est l'estimation et vice versa. Notamment, il faut être très prudent lors de l'utilisation des estimations dont le C.V. est très élevé (supérieur à 25 %), ce qui indique une très grande variabilité de l'estimation. Ces estimations ne devraient être utilisées qu'à titre indicatif. Le C.V. a également l'avantage de permettre de comparer la précision de différentes estimations entre elles. Le tableau suivant présente la relation entre le C.V. et la précision de l'estimation utilisée par l'ISQ.

Afin de permettre au lecteur de porter un jugement rapide sur la valeur d'une estimation, les coefficients de variation (C.V.) sont remplacés au sein des tableaux par des cotes de précision. Ces cotes sont définies dans le tableau de la page suivante.

Relation entre la valeur d'un coefficient de variation et la précision de l'estimation associée		
C.V.	Cote	Précision de l'estimation
≤ 5 %	A	excellente
5 % < C.V. ≤ 10 %	B	très bonne
10 % < C.V. ≤ 15 %	C	bonne
15 % < C.V. ≤ 25 %	D	passable
> 25 %	E	faible

² William G. Cochran, *Sampling techniques*, Harvard University, Wiley & sons, 1977 p. 160-161

³ William G. Cochran, *Sampling techniques*, Harvard University, Wiley & sons, 1977 p. 162

Dans le cadre de la présente enquête, les estimations pour le domaine des établissements provenant du secteur SCIAN 54 et celles pour le domaine regroupant les SCIAN 334, 3254 et 3364 affichent une bonne précision. En effet, on retrouve beaucoup de cotes A, B ou C pour les estimations de ces domaines. Cette performance peut s'expliquer par le fait qu'une grande proportion d'établissements de ces secteurs a été échantillonnée. Cependant, dans certaines catégories professionnelles, notamment les métiers et le personnel de la vente et des services et des services non spécialisés, les résultats obtenus cette année affichent une moins bonne précision que ceux de l'an passé. Une très grande variabilité dans les réponses obtenues semble la seule cause de cette perte de précision. Cette croissance des marges d'erreur se répercute également dans certains regroupements de ces catégories professionnelles. Une légère perte de précisions affecte donc les totaux de plusieurs catégories professionnelles. La plupart des estimations de totaux et de quotients effectuées pour l'ensemble des établissements sont malgré tout affectées d'une cote de valeur A, B ou C, ce qui signifie qu'elles sont assez précises et qu'elles peuvent être utilisées sans crainte. Toutefois, lorsque l'estimation d'un quotient ou d'une proportion est près de zéro, la précision de celle-ci tend à être moins satisfaisante (cotes de valeur D ou E). Puisque les tailles d'échantillon des autres secteurs de fabrication et de services ainsi que le secteur de la construction sont assez grandes, ce phénomène n'est toutefois pas trop présent ni pour ces secteurs ni pour l'ensemble des secteurs. En effet, des quotients et des proportions relativement près de zéro peuvent être estimés de façon assez précise même pour ces secteurs.

Il faut noter que les estimations faites à partir de certains domaines sont moins précises (cotes de valeur D ou E). Ce fait n'est pas surprenant étant donné la taille de l'échantillon disponible pour ces domaines. De tels cas se produisent notamment pour les questions 5, 6 et 8 puisque l'on retrouve peu d'établissements qui ont des postes vacants ou qui ont embauché des employés de l'extérieur du Québec. Ainsi, pour ces questions et spécialement pour la question 8, il est très hasardeux de faire des comparaisons entre les résultats des diverses réalisations de l'enquête puisque les marges d'erreur sont très grandes (parfois même plus grande que l'estimation elle-même). Des estimations affectées d'une cote de valeur D ou E devraient donc être utilisées avec la plus grande précaution et ne devraient pas être reprises hors contexte si elles ne sont pas accompagnées de leur marge d'erreur. Cette année, en raison de données plus variables, la même mise en garde peut s'appliquer aux questions 2 et 4 même si les domaines sont plus grands.

Pour les questions 4b et 8b, quelques établissements ont répondu qu'ils ne savaient pas combien il y avait eu de départs ou d'embauches à l'extérieur du Québec. Lorsque l'on calcule l'estimation associée à ces questions, on ne peut tenir compte de la réponse de ces établissements. Ceci engendre donc une légère sous-estimation pour les totaux produits. Cette sous-estimation est cependant faible car cette situation se produit pour une douzaine d'établissements. La même situation se produit à la question 6b. L'impact pour cette question est probablement légèrement supérieur puisque certains établissements ayant déclaré beaucoup de postes vacants à la question 5b ne savaient pas combien ils étaient depuis au moins quatre mois. Or, sur un grand nombre de postes vacants, on peut s'attendre à ce qu'un certain nombre le soient depuis un certain temps.

Lors de l'enquête 2003, certains gros établissements avaient répondu qu'ils n'avaient pas de poste vacant à la question 5. Or cette année, certains de ceux-ci ont déclaré un très grand nombre de postes vacants. Ceci nous amène à douter de la validité de la donnée de l'année passée pour ces quelques établissements. Ainsi, ils ont peut-être répondu ne pas avoir de poste vacant plutôt que dire qu'ils ne savaient pas combien ils en avaient. À la question 5b, on perçoit donc une hausse du

nombre de postes vacants entre les deux occurrences de l'enquête, mais elle n'est peut-être pas aussi grande que ne le laisse croire les données.

Par ailleurs, lorsque l'on compare les estimations du nombre total d'employés entre les enquêtes 2003 et 2004, on remarque qu'elles sont plus élevées au global cette année, bien que les marges d'erreur ne permettent pas de détecter de différence significative dans la majorité des cas. On remarque toutefois que cette hausse est principalement marquée pour les métiers et le personnel de la vente alors que certaines catégories professionnelles présentent une diminution du nombre d'employés. Par ailleurs, on remarque qu'une hausse importante du nombre d'employés s'est produite dans le secteur de la fabrication de produits aérospatiaux et de leurs pièces.

Soulignons que pour les estimations de proportions, la façon de les calculer en utilisant la variable indicatrice et un estimateur quotient (voir l'annexe C) ne nous permet pas d'obtenir des intervalles de confiance binomiaux pour des petites proportions. Ainsi, il est possible de voir des intervalles de confiance inférieurs à 0 ou supérieurs à 100. Il faut être prudent avec de telles estimations.

En ce qui concerne les indicateurs, plusieurs cotes de valeur D ou E sont obtenues. Habituellement, lorsque l'on obtient un C.V. très élevé pour une estimation, on est porté à ne pas faire confiance à l'estimation. Dans ce cas-ci, il faut savoir que l'indicateur est une différence de deux quantités. Dans une telle situation, il est vrai d'affirmer que la variabilité de l'estimation est très grande par rapport à l'estimation. Malgré cela, il ne faut pas penser qu'aucune conclusion ne peut être tirée. Par exemple, on pourrait, par le biais d'un test, montrer que l'estimation de la différence est ou n'est pas significativement différente de 0. D'autre part, il faut noter que lorsqu'une estimation est près de 0, le C.V. aura tendance à être très grand et instable. Dans une telle situation, le C.V. doit être utilisé avec précaution ou ne pas être utilisé du tout. Il est à noter que l'estimation d'une différence peut amener plusieurs cas où l'estimation est près de 0.

Finalement, rappelons que des estimations régionales sont produites pour les régions Montréal, centrales et ressources. La qualité des estimations produites pour ces régions est moins bonne que celle obtenue au niveau provincial. Des cotes de valeur A sont rarement obtenues pour ces régions. Les cotes de moins bonne qualité sont obtenues pour les régions centrales et ressources en raison de leur taille d'échantillon qui est beaucoup plus petite que celle de la région de Montréal. D'ailleurs, pour les régions ressources, il pourrait s'avérer très risqué de faire des comparaisons entre les estimations des diverses années d'enquêtes puisque ces dernières varient énormément d'une fois à l'autre en raison de la faible précision.

8.2 Les erreurs non dues à l'échantillonnage

Tel que mentionné précédemment, les erreurs non dues à l'échantillonnage peuvent provenir de plusieurs sources : un mauvais taux de réponse (global ou partiel), des questions difficiles à interpréter, des erreurs de saisie, des erreurs d'interprétation de l'interviewer ou des réponses fausses données (volontairement ou non) par les répondants.

Le taux de réponse global et la non-réponse partielle

Le taux de réponse global à une enquête est un indicateur de la qualité de celle-ci, notamment en regard des biais dus à la non-réponse qui pourraient être introduits dans les estimations. En effet,

plus la non-réponse est élevée, plus le risque est présent que des biais soient introduits dans les estimations en utilisant les données recueillies auprès des répondants pour inférer à l'ensemble de la population, les non-répondants pouvant présenter des caractéristiques différentes de celles des répondants. L'ISQ, comme tout organisme statistique, veille à limiter les risques de biais dans les enquêtes; c'est pourquoi il cherche à s'assurer du meilleur taux de réponse possible dans le cadre de toutes ses enquêtes, allant même jusqu'à s'abstenir de procéder à l'inférence statistique lorsque ce taux n'atteint pas un niveau satisfaisant. Dans un tel cas, des analyses de biais sont entreprises afin de mesurer les risques d'erreur et des correctifs sont apportés avant de produire les estimations.

Dans le cas présent, un très bon taux de réponse global (81,7 %) à l'enquête a été obtenu. De plus, les taux de réponse par strate (voir **tableau 3**) sont aussi satisfaisants puisqu'ils sont tous supérieurs à 75 %. Ces taux nous permettent d'être relativement rassurés quant aux possibilités de biais potentiels.

Pour ce qui est de la non-réponse partielle, 108 cas ont été rencontrés. Ce nombre est relativement faible, bien que supérieur à celui des enquêtes précédentes, ce qui nous permet d'être optimistes par rapport à la non-existence d'un biais quelconque.

La base de sondage

La base de sondage soulève un problème. En effet, si on peut supposer que celle-ci contient la presque totalité des établissements qui étaient en activité au moment de la mise à jour (juin 2004, dans le cas présent), il est certain qu'elle contient aussi des établissements qui ont cessé leurs activités. Ce phénomène est connu sous le nom de surcouverture de la population étudiée par la base de sondage.

Cette surcouverture implique qu'il est difficile de connaître la taille exacte de la population étudiée. Bien entendu, on peut faire certaines hypothèses : dans certaines enquêtes, le mode de fonctionnement fait en sorte qu'il est possible d'identifier toutes les unités échantillonnées qui ne sont pas admissibles et de projeter ce résultat à l'ensemble de la population étudiée. Ainsi, on estime la surcouverture au niveau de l'échantillon et on la suppose « représentative » de la surcouverture au niveau de la population. Malheureusement, la présente enquête n'appartient pas à cette catégorie.

En effet, dans le cadre de cette enquête, il n'a pas été possible d'identifier le statut de tous les établissements. Dans bien des cas de non-réponse, il n'était pas facile d'attribuer le statut d'admissibilité ou non à l'enquête. Par exemple, voici quelques cas de non-réponse où le statut était ambigu : pas de réponse au numéro composé, mauvais numéro de téléphone, ligne occupée. Pour de tels cas, il fallait donc poser l'hypothèse suivante : parmi ces cas, on retrouve un certain nombre de cas admissibles et un certain nombre de cas inadmissibles. Ces nombres ont pu être estimés, par strate, à l'aide de la répartition des établissements dont le statut était connu. En déterminant un nombre d'admissibles parmi les non-répondants, il a été possible de donner un poids aux répondants admissibles pour que ceux-ci représentent tous les admissibles de l'échantillon. De cette façon, une estimation de la population admissible à l'enquête a pu être produite.

L'interprétation des questions

En général, la majorité des répondants ont semblé bien comprendre les questions posées et les réponses fournies étaient pertinentes.

Les erreurs de saisie

La collecte des données a été réalisée à l'aide d'un logiciel de type ITAO (interview téléphonique assistée par ordinateur). Ce logiciel permet d'effectuer les sauts de questions de façon automatique et fait en sorte que les répondants admissibles à répondre à une question puissent y répondre. L'interviewer devait valider la réponse saisie pour chaque question avant de passer à la question suivante. Ainsi, lors de l'entrevue, trois types de validation apparaissent pour confirmer certaines données avec le répondant. Dans un premier temps, des questions de validation apparaissent à l'écran pour qu'il y ait une confirmation des totaux par le répondant (Questions 1, 2, 4, 5, 6, 8), puis, un second type de validation permet de vérifier la cohérence entre les questions. Une telle validation apparaît notamment à la question 6 afin que les réponses soient cohérentes avec celles de la question 5. Enfin, des bornes de validation ont été utilisées pour comparer le nombre d'employés de l'établissement dans chaque catégorie avec celui donné par le répondant l'année dernière. Ainsi, lorsqu'un écart important était observé, une confirmation était demandée au répondant. Par ailleurs, une nouvelle forme de validation historique a été instaurée cette année. Celle-ci se faisait après l'entrevue à l'aide d'une validation en lot et consistait à comparer le nombre d'employés au 31 décembre 2004, auquel on ajoute les départs en 2004 et on soustrait les embauches en 2004, avec le nombre d'employés au 31 décembre 2003. Lorsque la différence entre ces nombres d'employés était jugée trop grande selon les bornes établies, le répondant était rappelé et les données étaient validées de nouveau. Au total, 173 établissements ont ainsi été rappelés parmi lesquels 169 ont été rejoints. Les données de cette année ont été corrigées par 75 établissements suite à cette validation alors que 94 ont indiqué que l'erreur provenait de la donnée de l'an passé.

Par ailleurs, un programme informatique a été conçu afin de valider, à la fin du processus de collecte, toutes les données ayant été recueillies. De telles validations ont permis de relever quelques réponses inhabituelles. Pour ces quelques cas, un rappel téléphonique était effectué afin de valider l'information recueillie.

Toutes ces précautions ont fait en sorte qu'il n'y a pas lieu de craindre, outre mesure, que des erreurs de saisie aient pu amener des biais dans les résultats.

Les réponses volontairement faussées

Il n'existe pas de moyen absolu de valider la véracité de toutes les réponses reçues dans le cadre de cette enquête. Cependant, dans le contexte présent, rien ne laisse supposer que des répondants aient délibérément fourni des réponses erronées.

9 La confidentialité des résultats

L'ISQ est tenu de protéger la confidentialité des renseignements fournis par les répondants. Dans le cas de cette enquête, les résultats produits sous formes de tableaux ont donc été examinés. Lors

d'un tel examen, il est possible de déceler deux types de divulgation. Le premier type se produit lorsque la cellule d'un tableau contient peu de répondants. Le second type se produit, pour sa part, lorsque qu'un petit nombre d'établissements contribuent à un grand pourcentage d'une estimation présentée dans le tableau.

L'examen de l'ensemble des tableaux produits pour cette enquête a permis de déterminer que le premier et le deuxième type de divulgation pouvaient se produire. Afin d'éliminer ces risques, des cellules de tableaux ont été supprimées. Il est à noter que la suppression de cellules est plus importante pour les secteurs de fabrication de haute technologie (produits informatiques et électroniques, produits pharmaceutiques et produits aérospatiaux) ainsi que pour les secteurs SCIAN 5413, 5415 et 5417.

De plus, tel que mentionné précédemment, des données régionales sont produites dans le cadre de cette enquête. Cette production pose un problème de confidentialité supplémentaire puisqu'il est plus facile d'identifier un établissement lorsque l'on sait de quelle région il vient. Pour cette raison, la suppression de cellules est plus importante au niveau régional qu'au niveau provincial. Elle est particulièrement importante dans les régions ressources et centrales puisque ces deux régions sont plus petites que celle de Montréal.

Depuis l'année 2003, des données sont produites pour le secteur de la construction. Lors de la vérification de la confidentialité, des problèmes ont été détectés pour ce secteur. Ces problèmes ne sont pas majeurs et sont comparables à ceux détectés pour d'autres secteurs comparables de l'enquête. Il n'y a pas de difficulté particulière à masquer les données de ce secteur. Cependant, lorsque l'on masque les données d'un secteur, on doit s'assurer que les données masquées ne puissent être déduites à partir des autres secteurs de l'enquête. Or, dans le cas présent, il serait possible de déduire les données du secteur de la construction à l'aide des secteurs suivants:

1. Total (incluant le secteur de la construction)
2. Autres secteurs de fabrication + construction

Afin de préserver le maximum d'informations pour ces secteurs et d'assurer qu'il ne serait pas possible de déduire les données de la construction, il a été décidé, pour ces deux secteurs, de ne pas masquer la valeur présentée dans une cellule mais plutôt de la modifier quelque peu. Cette modification est appliquée seulement dans le cas où il est possible de déduire de l'information masquée pour le secteur de la construction. La majorité du temps la modification appliquée à la quantité estimée est mineure.

En terminant, il est à noter que pour l'année 2004, le nombre de suppressions de cellules pour des raisons de confidentialité s'est maintenu au même niveau que l'année précédente. Rappelons que pour l'année 2003, une nouvelle règle de vérification de la confidentialité avait été utilisée. Cette règle avait permis de diminuer le nombre de suppressions de cellules par rapport à l'année précédente.

10 Appréciation globale

Dans le contexte et en tenant compte des mises en garde des sections précédentes, on peut conclure que le potentiel analytique des données de l'enquête est très bon et qu'il n'existe pas de

problème apparent à inférer les résultats à la population étudiée. Toutefois, il y a lieu de faire preuve de prudence dans l'usage des résultats qui sont accompagnés d'une forte marge d'erreur ou d'une cote élevée (D ou E).

ANNEXES

A) La méthode pour la rotation de l'échantillon

La technique utilisée pour procéder à la rotation de l'échantillon est appelée échantillonnage colloqué modifié (en anglais : « modified collocated sampling »). Cette technique doit être appliquée dans chacune des strates échantillonnées. La première étape de cette méthode consiste à diviser l'intervalle $[0,1]$ en N intervalles (N étant le nombre d'établissements de la population dans une strate). Une permutation aléatoire de ces N intervalles est ensuite effectuée. La première unité de la population se voit ensuite attribuer un nombre aléatoire dans le premier intervalle (après permutation), la deuxième unité de la population un nombre aléatoire dans le deuxième intervalle et ainsi de suite pour toutes les unités de la population.

L'exemple qui suit permet de mieux comprendre la méthode. Supposons que l'on ait une population de taille $N=10$ dans une strate. Il faut d'abord diviser l'intervalle $[0,1]$ en 10 intervalles de même longueur. Chaque intervalle aura donc une longueur de 0,1. Ensuite, il faut permuer aléatoirement ces intervalles. Supposons que l'on obtienne aléatoirement la permutation suivante : 3, 5, 10, 1, 8, 7, 2, 9, 6 et 4. Ainsi, la première unité de la population se verra attribuer un nombre aléatoire entre 0,2 et 0,3; la deuxième unité un nombre entre 0,4 et 0,5 et ainsi de suite. Le tableau ci-dessous résume bien cet exemple.

Unité de la population	Permutation aléatoire	Intervalle correspondant
1	3	0,2-0,3
2	5	0,4-0,5
3	10	0,9-1,0
4	1	0,0-0,1
5	8	0,7-0,8
6	7	0,6-0,7
7	2	0,1-0,2
8	9	0,8-0,9
9	6	0,5-0,6
10	4	0,3-0,4

Lors du premier tirage d'échantillon, si on veut échantillonner avec une fraction de sondage de 0,4, on sélectionne toutes les unités dans l'intervalle 0,0-0,4. Les unités sélectionnées sont alors les unités 4, 7, 1 et 10.

La première étape pour procéder à la rotation de l'échantillon, lors de la mise à jour de la base de sondage, est de tenir compte des établissements morts et des nouveaux établissements. Les établissements dits morts sont ceux qui étaient sur l'ancienne base de sondage et qui ne sont plus sur la nouvelle parce qu'ils ont fermé leurs portes entre temps. À l'inverse, les nouveaux établissements sont ceux qui sont seulement sur la nouvelle base de sondage parce qu'ils ont vu le jour après la parution de l'ancienne base de sondage. Un établissement qui a changé de strate est considéré comme un mort dans son ancienne strate et comme un nouveau dans sa nouvelle strate. Les nouveaux se voient assigner à un intervalle les uns à la suite des autres. Ils reçoivent ensuite un numéro aléatoire dans cet intervalle. Dans l'exemple, le premier nouveau se verra attribuer un numéro aléatoire dans l'intervalle 0,2-0,3, le deuxième dans l'intervalle 0,4-0,5 et ainsi de suite. Lors

de la mise à jour suivante de la base de sondage, on assignera les intervalles à partir de celui avec lequel on a terminé la fois précédente. Par exemple, si le dernier nouveau se voit assigner l'intervalle 0,6-0,7, alors le premier nouveau lors de la prochaine mise à jour de la base de sondage se verra assigner l'intervalle 0,1-0,2.

En ce qui concerne les morts, lorsqu'ils sont identifiés par la nouvelle base de sondage, ils sont simplement retirés de la population. Toutefois, s'ils sont identifiés par une source externe à la base de sondage, ils doivent être conservés dans la population.

Cette façon de tenir compte des naissances et des morts a pour effet que la fraction de sondage peut varier légèrement d'une fois à l'autre. En effet, il se peut qu'il n'y ait pas exactement le même nombre d'unités assignées à chaque intervalle puisque certaines ont pu mourir et être supprimées. Cette méthode nous conduit donc à des échantillons dont on ne peut contrôler parfaitement la taille.

Une fois les naissances et les morts traitées, il est possible de procéder à la rotation de l'échantillon. Pour ce faire, il suffit de déplacer la fenêtre de sélection de l'échantillon. Ainsi, si on veut un chevauchement de 50%, c'est-à-dire que 50% des établissements présents lors de la première occurrence de l'enquête doivent revenir lors de la seconde occurrence, on doit échantillonner les unités se trouvant dans l'intervalle 0,2-0,6. Dans cet exemple, les unités 4 et 7 seraient remplacées dans le second échantillon par les unités 2 et 9, alors que les unités 1 et 10 seraient rééchantillonnées pour une deuxième fois.

Cette méthode est donc très efficace pour contrôler le pourcentage de chevauchement. Elle permet aussi d'assurer les répondants du temps maximal qu'ils demeureront dans l'échantillon. Dans l'exemple précédent, il s'agit de deux occurrences de l'enquête (puisque le chevauchement est de 50 %).

B) Les indicateurs

Cette section traite des indicateurs qui ont été créés à l'aide des résultats de cette enquête. Tout d'abord, mentionnons que le CETECH désirait un indicateur semblable à celui produit dans l'enquête trimestrielle de Statistique Canada sur les perspectives du monde des affaires. Cependant, cet indicateur n'est pondéré que par le nombre d'employés. Puisque cet indicateur n'est pas pondéré par le nombre d'établissements que représente chaque unité dans la population, il ne tient donc pas compte du fait qu'il est produit à partir d'un échantillon d'établissements. De ceci peut découler que l'indicateur calculé soit biaisé puisqu'il ne représente qu'une partie des établissements et qu'il ne peut être inféré à l'ensemble de la population. De plus, une autre conséquence de cette absence de pondération est qu'aucune mesure de précision n'est associée à cet indicateur. Afin de mieux comprendre l'indicateur de Statistique Canada, voici un exemple de calcul de celui-ci à partir d'un échantillon fictif de 5 établissements :

Établissements	Nombre d'employés	Réponse donnée
1. Grand établissement #1	10 000 employés	a augmenté
2. Grand établissement #2	10 000 employés	est resté le même
3. Petit établissement #1	100 employés	a diminué
4. Petit établissement #2	200 employés	a diminué
5. Petit établissement #3	150 employés	a diminué

Le calcul de l'indicateur de Statistique Canada s'effectue de cette façon:

On calcule, tout d'abord, le nombre d'employés dans l'échantillon:

$$10000+10000+100+200+150 = 20450 \text{ (A)}$$

On calcule, ensuite, le nombre d'employés pour une réponse « a augmenté »:

$$10000 \text{ (B)}$$

Puis, on compte le nombre d'employés pour une réponse « a diminué »:

$$100+200+150 = 450 \text{ (C)}$$

On peut maintenant calculer le pourcentage de réponses « a augmenté »:

$$B/A = 10000/20450 = 48,9 \% \text{ (D)}$$

Et le pourcentage de réponses « a diminué »:

$$C/A = 450/20450 = 2,2 \% \text{ (E)}$$

Le bilan d'opinion est alors:

$$D-E = 48,9 \% - 2,2 \% = 46,7 \% \text{ ou } +46,7$$

Bien souvent, dans les enquêtes entreprises, les plus grands établissements sont recensés et les plus petits sont échantillonnés. Un exemple d'une telle situation serait, dans l'exemple précédent, de fixer le nombre de grands établissements dans la population à 2 et le nombre de petits à 100. Les deux grands établissements sont recensés et alors que seulement trois petits établissements sur 100 sont échantillonnés. Dans cet exemple fictif, ne faire aucune pondération par le nombre d'établissements revient à ne pas tenir compte dans l'indicateur des 97 petits établissements qui n'ont pas été sélectionnés. Le fait de tenir ou de ne pas tenir compte de ces établissements peut avoir une grande influence sur l'indicateur. En effet, dans cet exemple, si on effectuait une pondération supplémentaire par le nombre d'établissements on obtiendrait **un bilan d'opinion négatif**.

Pour les raisons énumérées précédemment, l'ISQ a proposé un autre indicateur pour l'enquête sur le recrutement et l'emploi au Québec. Ce nouvel indicateur n'est finalement que l'indicateur de Statistique Canada auquel on ajoute une pondération par le nombre d'établissements. Cet indicateur a été accepté par le CETECH.

Afin de créer cet indicateur, on doit, tout d'abord, assigner une valeur (v) à la réponse donnée par le répondant i . Par exemple, à la question 9, on pourrait assigner la valeur 1 à la réponse « Meilleures que celles de 2004 », la valeur 0 à la réponse « Pas de changement par rapport à l'année 2004 » et la valeur -1 à la réponse « Moins bonnes que celles de 2004 ». Voici comment v est défini pour chacune des questions où un indicateur est construit:

Pour la question 9 :

$$v = \begin{cases} 1 & \text{si Meilleures que celles de 2004} \\ 0 & \text{si Pas de changement par rapport à l'année 2004} \\ -1 & \text{si Moins bonnes que celles de 2004} \end{cases}$$

Pour la question 10 :

$$v = \begin{cases} 1 & \text{si Le nombre sera plus élevé} \\ 0 & \text{si Le nombre n'aura pas changé} \\ -1 & \text{si Le nombre sera moins élevé} \end{cases}$$

Une fois v déterminée, on peut calculer l'indicateur ($\hat{I}_D$), pour chacune des questions, de la façon suivante :

$$\hat{I}_D = \frac{\sum_{h \in D} \sum_{i=1}^{n_h} w_{hi} ne_{hi} v_{hi}}{\sum_{h \in D} \sum_{i=1}^{n_h} w_{hi} ne_{hi}}$$

où n_h est le nombre d'établissements répondants dans la strate h , w_{hi} est le poids de l'établissement i de la strate h , ne_{hi} est le nombre d'employés de l'établissement i de la strate h , v_{hi} est la valeur de la réponse donnée par l'établissement i de la strate h et D est le domaine pour lequel l'indicateur est calculé.

Des quotients permettant de décomposer l'indicateur en plusieurs parties pour les questions 9 et 10 ont également été produits dans le cadre de cette enquête. Par exemple, pour la question 9, l'indicateur se décompose en trois parties. La première est la proportion de réponses « Meilleures que celles de 2004 » par rapport à l'ensemble des réponses, la deuxième est la proportion de réponses « Pas de changement par rapport à l'année 2004 » et la dernière est la proportion de réponses « Moins bonnes que celles de 2004 ». Ces quotients sont calculés à l'aide de la même pondération que celle utilisée pour l'indicateur.

Lors de l'enquête sur le semestre de juillet à décembre 2001, un autre indicateur a été élaboré. Ce dernier ne tient pas compte du nombre d'employés dans chaque établissement. Cet indicateur ($\hat{I}_{2D}$) sera présenté à la fin de l'annexe C qui traite du calcul des proportions d'établissements.

Il est important de noter que les deux indicateurs se présentent sous la forme de quotients. La marge d'erreur d'un quotient pouvant être calculée par les logiciels SAS et SUDAAN, une mesure de précision peut donc être obtenue pour chacun de ces indicateurs.

C) La méthode pour le calcul des proportions d'établissements

Puisque le fichier de travail contient 6 732 observations provenant des 6 646 établissements échantillonnés, calculer une simple proportion ne serait donc plus une proportion d'établissements. Des modifications au processus d'estimation avaient donc dû être apportées lors de l'enquête précédente et conservées lors de la présente enquête. Il a d'abord fallu créer une variable indicatrice prenant la valeur 1 si l'établissement est présent dans une seule région, 1/2 s'il est présent dans deux régions et 1/3 s'il est présent dans les trois régions. En fait, cette variable est construite de telle sorte que si on la somme sur les 4 630 enregistrements répondants du fichier, on retrouve le nombre d'établissements répondants de l'échantillon, à savoir 4 544. D'autre part, le poids d'échantillonnage d'un établissement était répété pour tous les autres enregistrements se rapportant à lui. En utilisant des données fictives, voici à quoi ressemble le fichier de travail :

Enregistrement du fichier	Établissement	Région	Indicatrice	Poids
1	1	1	1	1,5
2	2	1	1/3	2,2
3	2	2	1/3	2,2
4	2	3	1/3	2,2
5	3	2	1	4
6	4	2	1/2	3,8
7	4	3	1/2	3,8

Les proportions d'établissements se calculent comme des ratios dont le numérateur représente le nombre de mentions de la caractéristique à l'étude et le dénominateur est la somme des variables indicatrices. Les proportions ont donc la forme suivante :

$$\hat{p}_D = \frac{\sum_{h \in D} \sum_{i=1}^{n_h} w_{hi} x_{hi} A_{hi}}{\sum_{h \in D} \sum_{i=1}^{n_h} w_{hi} A_{hi}}$$

où n_h est le nombre d'unités répondantes dans la strate h , w_{hi} est le poids de l'unité i de la strate h , A_{hi} est la variable indicatrice de l'unité i de la strate h , x_{hi} est la valeur de la réponse donnée par l'unité i de la strate h et D est le domaine pour lequel la proportion est calculée.

Par exemple, pour la question 9, la proportion d'établissements répondants « Meilleures que celles de 2004 » peut être calculée à l'aide des variables suivantes :

$$x = \begin{cases} 1 & \text{si Meilleures que celles de 2004} \\ 0 & \text{sinon} \end{cases}$$

$$A = \begin{cases} 1 & \text{si l'établissement est présent dans une seule région} \\ 1/2 & \text{si l'établissement est présent dans deux régions} \\ 1/3 & \text{si l'établissement est présent dans trois régions} \end{cases} .$$

Pour faire des estimations par région, il s'agit de définir le domaine de telle sorte que seules les unités présentes dans cette région sont prises en compte dans le calcul.

L'indicateur ne tenant pas compte du nombre d'employés dans les établissements dont il est question à la fin de l'annexe B peut donc être calculé, pour chacune des questions, de la façon suivante :

$$\hat{I}_{2D} = \frac{\sum_{h \in D} \sum_{i=1}^{n_h} w_{hi} v_{hi} A_{hi}}{\sum_{h \in D} \sum_{i=1}^{n_h} w_{hi} A_{hi}} .$$

où v est donné à l'annexe B pour chacune des questions.