

**Enquête sur le recrutement
et l'emploi au Québec
(année 2005)**

Rapport méthodologique

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

Institut de la statistique du Québec

Juillet 2006

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	12
6 La collecte des données	14
7 L'estimation.....	19
8 La qualité des estimations	21
9 La confidentialité des résultats.....	27
10 Appréciation globale	28
Annexes.....	29
A) La méthode pour la rotation de l'échantillon.....	30
B) La méthode pour le calcul des proportions d'établissements	32

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 2005	10
Tableau 3 Résultats de la collecte de données par strate	16

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. Une des spécificités de cette enquête est le fait qu'elle permet de mesurer 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, la formation ainsi que les embauches et les départs d'employés 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 2005.

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.

Notons que les établissements des secteurs de services définis par les codes du système de classification des industries de l'Amérique du Nord (SCIAN) 813110, 813310, 813410 et 813910 ont été exclus de la population visée. 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.

3 La base de sondage

La base de sondage utilisée pour l'enquête est construite à partir de la Base de données du registre des entreprises (BDRE) de Statistique Canada mise à jour en juin 2005 de laquelle on extrait les établissements appartenant à la population visée par l'enquête. Il y a eu un changement de base de sondage par rapport à l'enquête portant sur l'année 2004 puisque c'était la version de juin 2004 de la BDRE qui avait alors été utilisée. 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'emplacement 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. La BDRE contient également un numéro de téléphone. 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, 122 lettres ont été expédiées et 46 d'entre elles, soit près de 38 %, ont rappelé ou ont retourné le formulaire nous fournissant un numéro de téléphone permettant de les rejoindre. Cette stratégie, implantée l'année dernière, permet de bonifier la base de sondage et de 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. L'inclusion du niveau emplacement dans la base de sondage 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 2004 et 2005, la validation faite en 2004 a été conservée pour 2005 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. Lors de l'édition 2004 de l'enquête, seulement sept secteurs d'activités étaient retenus pour la stratification. Ce nombre est passé à 19 pour l'édition 2005 en raison principalement de la désagrégation des secteurs de fabrication et de services. Notons qu'à l'instar des enquêtes précédentes, c'est le SCIAN 2002 qui a été utilisé dans cette enquête. Les secteurs retenus se définissent comme suit :

- Construction (code du SCIAN : 23)
- Conception de systèmes informatiques (code du SCIAN : 5415)
- Architecture, génie et services connexes (code du SCIAN : 5413)
- Services de recherche et de développement scientifiques (code du SCIAN : 5417)
- Fabrication tertiaire à forte intensité de main d'œuvre (codes du SCIAN : 311-316, 337, 339)

- Fabrication primaire (codes du SCIAN : 321, 322, 324, 327,331)
- Fabrication secondaire (codes du SCIAN 3251-3253, 3256-3259, 326, 332)
- Fabrication de produits pharmaceutiques et de médicaments (code du SCIAN : 3254)
- Fabrication tertiaire à forte intensité de capital (codes du SCIAN : 323, 333, 335, 3361-3363, 3365-3369)
- Fabrication de produits informatiques et électroniques (code du SCIAN : 334)
- Fabrication de produits aérospatiaux et de leurs pièces (code du SCIAN : 3364)
- Commerce (codes du SCIAN : 41, 44-45)
- Transport et entreposage (codes du SCIAN : 48-49)
- Finance, assurance, immobilier et location (codes du SCIAN : 52-53)
- Services professionnels (codes du SCIAN : 5411, 5412, 5414, 5416, 5418, 5419)
- Gestion d'entreprises, services administratifs et autres services de soutien (codes du SCIAN : 55-56)
- Information, culture et loisirs (incluant arts et spectacles) (codes du SCIAN : 51, 71)
- Hébergement et restauration (code du SCIAN : 72)
- Autres services (code du SCIAN : 81).

À l'intérieur de chacun de ces secteurs, on a d'abord 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.

Pour éviter de devoir refaire ce travail chaque année et de créer de trop nombreux changements dans les strates, ce travail a été fait lors de l'élaboration du plan de sondage de l'enquête 2005 (à partir des données de la BDRE de juin 2005) et les points de coupures des strates en termes d'effectifs seront préservés pour les enquêtes de 2006 et 2007.

Par le passé, il n'y avait que deux strates de tailles définies de cette façon. Toutefois, il y avait lieu de se questionner sur la possibilité de biais qui pouvaient découler d'une telle approche. Ainsi, afin notamment de réduire ce biais potentiel, il a été décidé de recenser les plus grands établissements de chacun des secteurs. En effet, ces établissements, s'ils ne sont pas échantillonnés, ne sont souvent pas adéquatement représentés par d'autres et l'estimation sans ces derniers peut être affectée. La strate des grands établissements a donc été divisée de nouveau en deux strates : l'une recensée et l'autre échantillonnée. La méthode de Lavallée-Hidiroglou a été utilisée pour établir la coupure entre ces deux strates¹. Cette méthode consiste essentiellement à diviser les établissements entre les strates recensées et échantillonnées de manière à minimiser la taille de l'échantillon pour un niveau de précision donné.

¹ Lavallée P. et M.A. Hidiroglou. 1998. Sur la stratification de populations asymétriques. *Techniques d'enquête*. 14(1) : 35-45.

En définitive, pour chacun des secteurs, les établissements sont divisés en trois strates de taille : l'une contenant les grands établissements (partie recensée de la strate des grands), l'autre les moyens (partie échantillonnée de la strate des grands) et la troisième les petits. Les secteurs de la conception de systèmes informatiques, d'architecture, génie et services connexes, de services de recherche et de développement ainsi que des services professionnels ne sont toutefois divisés qu'en deux strates : grands et petits établissements alors que les secteurs de fabrication de haute technologie, qui sont recensées, ne sont pas divisés par taille.

Au total, on dispose donc de 47 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.

4.2 La taille et la répartition de l'échantillon

Dans l'entente avec le CETECH, une taille d'échantillon de 6 692 établissements était visée pour cette enquête. 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 6 773 employeurs québécois. Lors des précédentes enquêtes, les analyses portaient sur un plus grand nombre d'établissements que ceux enquêtés sur le terrain. Ceci était dû aux établissements déclarés inadmissibles de façon permanente lors des réalisations précédentes de l'enquête qui n'étaient pas enquêtés à nouveau. Or, cette année, il a été décidé de valider à nouveau cette liste afin de confirmer leur statut d'inadmissible. Ceci est une précaution à prendre après quelques années. L'an prochain, la procédure habituelle consistant à conserver ces établissements seulement pour l'analyse en les considérant d'emblée comme inadmissibles sera à nouveau utilisée.

La taille et la répartition de l'échantillon d'établissements retenus pour l'année 2005 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, les moyens 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 (%)
Construction	Grands	190	190	100,0	26,9
	Moyens	1 513	404	26,7	40,1
	Petits	5 678	334	5,9	33,0

Conception de systèmes informatiques	Grands	62	62	100,0	57,5
	Petits	931	269	28,9	42,5
Architecture, génie et services connexes	Grands	115	115	100,0	56,8
	Petits	867	226	26,1	43,2
Services de recherche et de développement scientifiques	Grands	42	42	100,0	59,4
	Petits	300	145	48,3	40,6
Fabrication tertiaire à forte intensité de main-d'œuvre	Grands	27	27	100,0	13,8
	Moyens	428	83	19,4	47,3
	Petits	2 994	83	2,8	38,9
Fabrication "primaire"	Grands	34	34	100,0	25,2
	Moyens	281	68	24,2	45,0
	Petits	1 312	78	5,9	29,8
Fabrication "secondaire"	Grands	12	12	100,0	10,6
	Moyens	204	46	22,5	39,1
	Petits	1 838	93	5,1	50,3
Fabrication de produits pharmaceutiques et de médicaments	Toutes les Tailles	65	65	100,0	100,0
Fabrication tertiaire à forte intensité de capital	Grands	14	14	100,0	19,2
	Moyens	183	52	28,4	40,5
	Petits	1 537	58	3,8	40,3
Fabrication de produits informatiques et électroniques	Toutes les Tailles	301	301	100,0	100,0
Fabrication de produits aérospatiaux et de leurs pièces	Toutes les Tailles	65	65	100,0	100,0
Commerce	Grands	31	31	100,0	5,0
	Moyens	2 356	301	12,8	46,4
	Petits	20 750	216	1,0	48,6
Transport et entreposage	Grands	84	84	100,0	48,1
	Moyens	88	43	48,9	9,0
	Petits	3 174	284	8,9	42,9
Finance, assurances, immobilier et location	Grands	225	225	100,0	43,9
	Moyens	321	103	32,1	11,5
	Petits	5 603	336	6,0	44,6
Services professionnels seulement	Grands	242	242	100,0	57,5
	Petits	2 823	376	13,3	42,5

Tableau 1(suite)

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 (%)
Gestion d'entreprises, services administratifs et autres services de soutien	Grands	108	108	100,0	35,7
	Moyens	505	96	19,0	27,9
	Petits	4 182	260	6,2	36,4

Information, culture et loisirs (incluant arts et spectacles)	Grands	36	36	100,0	26,3
	Moyens	171	40	23,4	26,5
	Petits	3 092	248	8,0	47,2
Hébergement et restauration	Grands	64	64	100,0	10,5
	Moyens	3 265	83	2,5	71,9
	Petits	7 019	77	1,1	17,6
Autres services	Grands	172	172	100,0	32,0
	Moyens	533	218	40,9	23,6
	Petits	3 714	264	7,1	44,4
Total		77 551	6 773	8,7	

Habituellement, une rotation de l'échantillon a lieu entre les différentes éditions de l'enquête. 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. Toutefois, cette année, en raison de la désagrégation des secteurs de fabrication et de services ainsi que du recensement des grands établissements, un très grand nombre de nouveaux établissements entraient dans l'échantillon. Il a donc été décidé de ne faire aucune rotation des établissements dans les strates échantillonnées, et ce afin de préserver un bon chevauchement entre les échantillons des années 2004 et 2005. Les établissements morts ont toutefois été remplacés par d'autres afin de maintenir la taille d'échantillon. Toutefois, comme la rotation de l'échantillon devrait être à nouveau appliquée lors d'une prochaine édition de l'enquête, le tirage de l'échantillon sera tout de même fait à l'aide d'une méthode permettant une telle rotation. D'ailleurs, plusieurs méthodes existent pour faire la rotation d'un échantillon. Celle qui est utilisée habituellement dans cette enquête est appelée échantillonnage colloqué modifié² (en anglais : « modified collocated sampling »). Cette méthode a donc été appliquée pour tirer l'échantillon dans chacune des 28 strates échantillonnées. Une description détaillée et un exemple de cette méthode se trouvent à l'annexe A.

Le **tableau 2** présente le renouvellement de l'échantillon pour l'enquête de 2005. En effet, on y retrouve la répartition des établissements échantillonnés pour l'enquête de 2005 selon leur statut à l'enquête de 2004. On s'aperçoit, en regardant ce tableau, que 2 953 établissements n'ayant pas été échantillonnés lors de l'enquête annuelle 2004 l'ont été lors de l'enquête annuelle 2005. Ces nouveaux établissements sont entrés dans l'échantillon en raison des modifications au plan de sondage 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 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.

² K.P. Srinath et Ronald M. Carpenter, *Business Survey Methods*, Wiley & sons, 1995, chapitre 10, pp. 171-183

Tableau 2
Renouvellement de l'échantillon pour l'enquête annuelle de 2005

Secteurs d'activité	Taille	Échantillonnés pour l'enquête actuelle ¹ (n)	Échantillonnés pour l'enquête précédente ^{2,3} (n)	Non échantillonnés pour l'enquête précédente (n)
Construction	Grands	190	82	108
	Moyens	404	308	96
	Petits	334	243	91
Conception de systèmes informatiques	Grands	62	51	11
	Petits	269	206	63
Architecture, génie et services connexes	Grands	115	98	17
	Petits	226	160	66
Services de recherche et de développement scientifiques	Grands	42	39	3
	Petits	145	80	65
Fabrication tertiaire à forte intensité de main-d'œuvre	Grands	27	6	21
	Moyens	83	81	2
	Petits	83	73	10
Fabrication "primaire"	Grands	34	7	27
	Moyens	68	67	1
	Petits	78	71	7
Fabrication "secondaire"	Grands	12	2	10
	Moyens	46	42	4
	Petits	93	81	12
Fabrication de produits pharmaceutiques et de médicaments	Toutes les Tailles	65	59	6
Fabrication tertiaire à forte intensité de capital	Grands	14	6	8
	Moyens	52	49	3
	Petits	58	54	4
Fabrication de produits informatiques et électroniques	Toutes les Tailles	301	254	47
Fabrication de produits aérospatiaux et de leurs pièces	Toutes les Tailles	65	62	3
Commerce	Grands	31	6	25
	Moyens	301	267	34
	Petits	216	183	33
Transport et entreposage	Grands	84	8	76
	Moyens	43	11	32
	Petits	284	84	200
Finance, assurances, immobilier et location	Grands	225	38	187
	Moyens	103	35	68
	Petits	336	86	250
Services professionnels seulement	Grands	242	189	53
	Petits	376	285	91

Tableau 2 (suite)

Renouvellement de l'échantillon pour l'enquête annuelle de 2005

Secteurs d'activité	Taille	Échantillonnés	Échantillonnés	Non échantillonnés
		pour l'enquête actuelle ¹ (n)	pour l'enquête précédente ^{2,3} (n)	pour l'enquête précédente (n)
Gestion d'entreprises, services administratifs et autres services de soutien	Grands	108	11	97
	Moyens	96	46	50
	Petits	260	83	177
Information, culture et loisirs (incluant arts et spectacles)	Grands	36	3	33
	Moyens	40	18	22
	Petits	248	82	166
Hébergement et restauration	Grands	64	6	58
	Moyens	83	60	23
	Petits	77	57	20
Autres services	Grands	172	15	157
	Moyens	218	5	213
	Petits	264	61	203
Total		6 773	3 820	2 953

¹ L'enquête actuelle est l'enquête annuelle de 2005.

² L'enquête précédente est l'enquête annuelle de 2004.

³ Dans les rapports méthodologiques des précédentes enquêtes, les établissements échantillonnés l'année précédente étaient séparés en deux catégories à savoir les admissibles et inadmissibles temporaires et les inadmissibles permanents. Or, comme tous les établissements ont été enquêtés cette année peu importe leur statut à l'enquête de 2004, il n'y a pas lieu de faire une telle distinction.

On remarque que le nombre de nouveaux établissements est plus élevé cette année que l'an passé. Ceci s'explique principalement par les modifications importantes au plan de sondage. Ainsi, la désagrégation des secteurs des autres établissements de fabrication et de services a nécessité l'ajout de plusieurs établissements dans ces secteurs par rapport au nombre échantillonnés auparavant. De même, la décision de recenser les grands établissements de tous les secteurs d'activité a contribué à faire entrer des nouveaux établissements dans l'échantillon.

Enfin, la quatrième colonne du **tableau 2** montre qu'il y a 3 820 établissements échantillonnés lors de l'enquête 2004 qui sont à nouveau dans l'échantillon de l'enquête 2005, et ce peu importe le statut d'admissibilité qu'ils avaient en 2004.

De la population de 77 551 établissements, 6 773 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 de chacun des secteurs d'activité. À l'intérieur des autres strates, l'échantillonnage colloqué modifié a permis de sélectionner l'échantillon.

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. En fait, 328 établissements (parmi les 6 773 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 6 943 observations : 6 445 établissements ayant un seul emplacement, 127 ayant plusieurs emplacements tous concentrés dans la même région, 70 ayant des emplacements dans deux régions, 50 ayant des emplacements dans trois régions, 81 ayant plusieurs emplacements mais étant non-répondants ou inadmissibles (dans ce cas, une seule observation apparaît au fichier).

4.4 La répartition par taille de l'échantillon

Une nouveauté a été introduite à l'enquête cette année consistant à produire des estimations par taille d'établissements. Des estimations sont donc produites pour les tailles d'établissements suivantes :

- 5 à 20 employés
- 20 à 99 employés
- 100 employés et plus.

Selon la BDRE, 2 721 établissements de l'échantillon de 6 773 établissements avaient moins de 20 employés, 2 339 avaient de 20 et 99 employés et 1 713 avaient 100 employés ou plus.

5 Le questionnaire

Des modifications assez importantes ont été apportées au questionnaire pour la version 2005 de l'enquête. En effet, plusieurs questions présentes dans le passé ont été supprimées. C'est le cas notamment des questions sur les perspectives d'affaires, sur les difficultés de recrutement et des raisons de départs d'employés. De nouvelles questions se sont toutefois ajoutées, portant sur la formation des employés et sur la gestion des ressources humaines de l'établissement. Par ailleurs, suite à la collecte de 2004, le Centre d'étude sur l'emploi et la technologie (CETECH) et l'ISQ ont proposé quelques légères modifications aux questions communes aux deux enquêtes visant à faciliter le traitement des données et ont collaboré à la production de la version finale du questionnaire. Les résultats de quelques unes de ces modifications sur le terrain seront discutés à la section 6.1. La nouvelle version du questionnaire est composée de 13 questions plutôt que 11 comme c'était le cas en 2004. L'ensemble des thèmes qui sont abordés dans le questionnaire de l'enquête annuelle 2005 sont les suivants:

1. Nombre de personnes qui travaillaient dans l'établissement au 31 décembre 2005
2. Embauche de nouveaux employés entre le 1^{er} janvier et le 31 décembre 2005
3. Départs d'employés entre le 1^{er} janvier et le 31 décembre 2005

4. Postes actuellement vacants
5. Postes vacants depuis au moins quatre mois
6. Mesures prises pour faire face à la difficulté que posent ces postes demeurés longtemps vacants
7. Activités de formations offertes aux employés entre le 1^{er} janvier et le 31 décembre 2005
8. Présence d'un service du personnel ou d'une direction des ressources humaines dans l'établissement
9. Présence d'une personne s'occupant des ressources humaines
10. Statut de la personne s'occupant des ressources humaines
11. Pourcentage du temps de cette personne consacré aux ressources humaines
12. Présence ou absence de mesures de gestion touchant divers aspects (formation du personnel, processus d'embauche et de recrutement, roulement de personnel etc.) ainsi que nécessité d'améliorer ou non ces aspects.
13. Favorisation du coaching et du mentorat.

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 6 et 8 à 13 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).

Comme par les années passées, 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 3 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 les précédentes enquêtes, 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 ainsi que de s'assurer de l'adresse d'expédition et du nom du répondant. Il s'est déroulé entre le 16 janvier et le 17 février. Tout comme en 2004, un filtre a également été fait sur les 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. Dans le cas des établissements enquêtés en 2004 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 21 mars au 2 juin 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 aussi 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 contact 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 20 intervieweurs, sous la supervision constante d'une chargée d'enquête. Les intervieweurs 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. Cette validation 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 intervieweurs.

En moyenne, cinq 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 10 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.

Souignons qu'afin d'aider les répondants à répartir leurs employés entre les divers groupes de professions, une annexe détaillée a été utilisée par les intervieweurs. 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

il y a quelques années 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 répondants ayant soulevé la question à savoir s'ils devaient considérer comme «nouvel employé» une personne ayant déjà été à leur emploi, soit dans un autre emploi, soit dans le même emploi mais pas au cours de l'année. Puisqu'elle causait de la confusion, la mention «nouveaux employés» avait donc été retirée pour l'enquête 2004. Le terme «nouveau personnel» a toutefois été utilisé cette année soulevant les mêmes interrogations qu'en 2003. Il serait donc préférable de revenir au libellé de 2004.

Bien que le questionnaire soit en général bien compris par les répondants, certaines améliorations pourraient encore y être apportées. Ainsi, certaines définitions et directives s'appliquant à plusieurs questions, il serait souhaitable d'ajouter en page 2 du questionnaire deux sections portant sur les définitions et les directives. La première contiendrait les définitions des termes employé, emploi étudiant et catégories professionnelles. La seconde donnerait des directives sur les employés à inclure et à exclure ainsi que sur les mouvements inter-établissements. Elle contiendrait également des exemples permettant de préciser les employés engagés et mis à pieds durant la même année qui doivent être comptabilisés.

Dans le questionnaire informatisé, une mention était ajoutée précisant les raisons de départs que l'on souhaite inclure (démissions, retraites, mises à pied ou autres raisons). Cette précision devrait être ajoutée au questionnaire papier afin que les répondants incluent tous les départs d'employés. Par ailleurs, la question 7 a posé un réel problème. En effet, certains établissements ont de la difficulté à répondre à cette question puisque les informations centralisées aux ressources humaines étaient souvent colligées en fonction de besoins budgétaires. Plusieurs ont donc indiqué ne pouvoir fournir des informations que sur le nombre d'activités de formation qui avaient été financées et non sur le nombre de personnes ayant bénéficié de ces formations. Les conséquences de ce phénomène sur les estimations seront discutées à la section 8. Enfin, la question 12 a parfois posé problème puisque les informations demandées ne relevaient pas toujours de l'établissement, mais plutôt de l'entreprise. Quelques autres légères modifications pourraient être apportées au questionnaire et à l'annexe pour la prochaine édition de l'enquête.

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 et des établissements ayant plusieurs emplacements. 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 4 788 établissements ce qui donne un très bon taux de réponse de 84,9 % (81,7 % lors de l'enquête 2004). Ce taux est supérieur à tous ceux obtenus lors des précédentes éditions de l'enquête. Il est d'autant plus surprenant d'obtenir un tel taux de réponse puisque le nombre de nouveaux établissements est très élevé encore une fois

cette année. Le nombre d'enregistrements servant dans les calculs d'estimation est 4 958 (4 668 enregistrements provenant d'établissements présents dans une seule région, 140 provenant d'établissements présents dans deux régions et 150 provenant d'établissements présents dans trois régions).

Pour les établissements ayant un seul emplacement, la durée moyenne pour compléter une entrevue était de 11 minutes (elle était de 7 minutes en 2004), la durée maximale de 1 heure et 19 minutes et la durée minimale de 51 secondes. La durée maximale se produisait lors de situations plus complexes, comme par exemple pour les entreprises possédant plusieurs employés ou pour des fusions récentes. En raison d'une partie très importante du travail dans les dossiers d'établissements ayant plusieurs emplacements qui est faite soit à partir de liste papier ou encore de recherche pour bien cerner les entités échantillonnées, les données administratives ne peuvent être comptabilisées et refléter correctement le temps consacré à ces dossiers.

La collecte s'est soldée par un taux de collaboration de 98,1 %. L'année dernière, ce taux se situait à 98,3 %. 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 1 133 établissements (parmi les 6 773 contactés par les intervieweurs de l'ISQ) qui n'étaient pas admissibles à l'enquête : 717 où il y a moins de 5 employés, 349 fermés définitivement, 17 fermés temporairement, 5 du secteur public, 4 hors population et 38 en double. Soulignons que le nombre d'inadmissible est supérieur à celui de l'année 2004 principalement en raison du fait que tous les établissements de l'échantillon étaient enquêtés même s'ils avaient déjà été déclarés inadmissibles permanents lors d'une précédente enquête. Évidemment, la très grande majorité de ces cas se sont avérés inadmissibles une fois de plus.

Le **tableau 3** présente les résultats de la collecte pour chacune des strates de l'enquête. Dans ce tableau, on constate qu'uniquement deux taux de réponse par strate sont inférieurs à 70 % et seulement neuf sont inférieurs à 80 %. Trois strates ont même des taux de réponse supérieurs à 90 %, ce qui est remarquable. Les taux de réponse par strate varient entre 58,1 % et 91,7 %. En raison des importants changements dans les strates entre les années 2004 et 2005 de l'enquête, il est difficile de discuter l'évolution des taux de réponse dans chacune d'entre elles. On constate toutefois, dans le cas des strates de fabrication de haute technologie, une légère baisse des taux de réponse, bien que ces derniers demeurent très bons. Des efforts considérables sont mis pour obtenir de bons taux de réponse dans ces strates névralgiques pour l'enquête.

Tableau 3
Résultats de la collecte de données par strate

Secteurs d'activité	Taille	Échantillon (n)	Échantillon admissible (n)	Répondants (n)	Taux de réponse (%)
Construction	Grands	190	161	141	87,6

	Moyens	404	360	322	89,4
	Petits	334	276	239	86,6
Conception de systèmes informatiques	Grands	62	55	49	89,1
	Petits	269	217	181	83,4
Architecture, génie et services connexes	Grands	115	92	81	88,0
	Petits	226	181	162	89,5
Services de recherche et de développement scientifiques	Grands	42	38	28	73,7
	Petits	145	120	97	80,8
Fabrication tertiaire à forte intensité de main-d'œuvre	Grands	27	25	20	80,0
	Moyens	83	82	68	82,9
	Petits	83	74	64	86,5
Fabrication "primaire"	Grands	34	33	29	87,9
	Moyens	68	62	54	87,1
	Petits	78	72	66	91,7
Fabrication "secondaire"	Grands	12	12	10	83,3
	Moyens	46	44	38	86,4
	Petits	93	82	72	87,8
Fabrication de produits pharmaceutiques et de médicaments	Toutes les Tailles	65	58	47	81,0
Fabrication tertiaire à forte intensité de capital	Grands	14	14	11	78,6
	Moyens	52	52	43	82,7
	Petits	58	51	45	88,2
Fabrication de produits informatiques et électroniques	Toutes les Tailles	301	264	217	82,2
Fabrication de produits aérospatiaux et de leurs pièces	Toutes les Tailles	65	62	52	83,9
Commerce	Grands	31	29	24	82,8
	Moyens	301	281	246	87,5
	Petits	216	184	155	84,2
Transport et entreposage	Grands	84	74	66	89,2
	Moyens	43	36	27	75,0
	Petits	284	234	200	85,5
Finance, assurances, immobilier et location	Grands	225	169	143	84,6
	Moyens	103	85	77	90,6
	Petits	336	230	196	85,2
Services professionnels seulement	Grands	242	211	168	79,6
	Petits	376	265	227	85,7

Tableau 3 (suite)

Résultats de la collecte de données par strate

Secteurs d'activité	Taille	Échantillon (n)	Échantillon admissible (n)	Répondants (n)	Taux de réponse (%)
Gestion d'entreprises, services administratifs et autres services de soutien	Grands	108	92	63	68,5
	Moyens	96	76	65	85,5
	Petits	260	176	137	77,8
Information, culture et loisirs (incluant	Grands	36	31	18	58,1

arts et spectacles)	Moyens	40	34	29	85,3
	Petits	248	194	165	85,1
Hébergement et restauration	Grands	64	62	56	90,3
	Moyens	83	72	55	76,4
	Petits	77	64	51	79,7
Autres services	Grands	172	154	131	85,1
	Moyens	218	197	177	89,8
	Petits	264	203	176	86,7
Total		6 773	5 640	4 788	84,9

En ce qui a trait à la non-réponse partielle, seulement un petit nombre de cas ont été rencontrés. En effet, pour 157 établissements (3,3 % des répondants), des réponses manquaient à certaines questions. Voici les différents types d'imputation utilisées. Dans 125 cas, il a fallu procéder à une imputation à l'aide des informations de la BDRE (informations sur les entreprises, sur les établissements et sur les emplacements) pour répartir le nombre d'employés d'une entreprise entre ses divers établissements. Dans certains de ces cas, le répondant fournissait en surplus des informations pour des établissements qui n'avaient pas été enquêtés. De plus, 25 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 5 cas pour regrouper ces informations. Enfin, une imputation par donneur a été utilisée pour imputer 2 cas où de la non-réponse a été obtenue aux questions 4, 5, 6 et 7. Notons que le nombre de cas de non-réponse partielle a augmenté par rapport à l'enquête précédente alors qu'il y en avait eu 107 cas (2,4 % des répondants). Cette augmentation est due en partie au plus grand nombre de grands établissements recensés. Ces derniers sont plus susceptibles d'être imputés.

On remarque que pour le secteur de fabrication de produits aérospatiaux et de leurs pièces, près du tiers (30,8 %) des établissements répondants ont été imputés ce qui correspond à 16 cas sur les 52 établissements répondants. Il y a donc une augmentation de l'imputation pour ce secteur puisque, l'an dernier, l'imputation avait été utilisée pour seulement 22,9 % (11 cas sur 48 répondants) des établissements répondants. Par ailleurs, pour la première fois à l'enquête, un autre secteur affiche un taux d'imputation supérieur à 10 %. Il s'agit du secteur de l'architecture, génie et services connexes avec 26 cas imputés sur les 243 répondants (10,7 %). L'an dernier, le taux d'imputation pour ce secteur se chiffrait plutôt à 7,2 %. Pour la majorité des autres secteurs, 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 se retrouve principalement 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. 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 qu'il nous faut 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 le secteur de fabrication de produits aérospatiaux et de leurs pièces ainsi que pour le secteur de l'architecture, génie et services connexes 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 exemple, 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 4 788 répondants admissibles, 1 133 répondants inadmissibles et 852 non-répondants. Seuls ces derniers possèdent un statut inconnu. Pour cette raison, l'hypothèse suivante a été utilisée : parmi les 852 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 les strates des moyens et petits établissements du secteur de la construction, dans la strate des petits établissements d'architecture, génie et services connexes, dans la strate des grands établissements de finance, assurance, immobilier et location ainsi que dans la strate des petits établissements des services professionnels, 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 cinq strates pour les établissements de la région de Montréal sont de 86 %, 82 %, 87 %, 80 % et 83 % tandis que ceux pour les établissements des autres régions sont respectivement de 95 %, 93 %, 96 %, 100 % et 97 %. À l'inverse, dans la strate des grands établissements de la construction, le taux de réponse est plus élevé parmi les établissements de la région de Montréal que parmi les établissements des autres régions : 92 % versus 79 %. En conséquence, l'ajustement de non-réponse dans toutes ces strates diffère selon la provenance de l'établissement.

L'ajustement de non-réponse étant fait à l'intérieur de chacune des strates, la désagrégation des anciennes strates des autres secteurs de fabrication et de services, a eu pour effet d'augmenter le nombre de strates et par le fait même de diminuer le nombre d'établissements présents dans chacune des strates. Ainsi, lors des précédentes enquêtes, des différences dans la probabilité de répondre selon la provenance des établissements étaient identifiées dans les strates des autres

secteurs de fabrication et de services. Or, en découpant ces secteurs, cette différence est toujours visible dans les secteurs plus fins, mais elle n'est plus significative statistiquement. Ainsi, aucun ajustement de non-réponse n'est fait dans la majorité des strates des autres secteurs de services et de fabrication. La conséquence de ceci est une augmentation des poids des régions centrales et ressources et une diminution de ceux de la région de Montréal par rapport à ce qu'ils étaient l'année dernière (lors de l'ajustement pour la non-réponse, les établissements des régions qui répondent davantage ont un poids plus faible que ceux des régions répondant moins bien). Cette différence dans les poids entraîne inévitablement une différence dans les estimations régionales rendant les comparaisons entre les estimations régionales de 2004 et 2005 impossibles. Ce sujet sera discuté davantage à la section 8.1.

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 les strates des petits établissements des services professionnels, scientifiques et techniques ainsi que dans les strates des petits et moyens établissements de tous les autres secteurs. Pour ce qui est des autres strates (celles des secteurs de fabrication de haute technologie et celles des grands établissements), un recensement a été effectué.

Pour cette étude, les estimations suivantes sont produites :

- totaux pour les questions 1, 2b, 3b, 4b, 5b et 7b
- proportions d'établissements pour les questions 2a, 3a, 4a, 5a, 5b, 6, 7a, 8, 9, 10, 11, 12 et 13
- quotients pour les questions 2a, 3a, 4a, 5a, 7a, 8, 9, 10, 11, 12 et 13.

En raison de caractère exclusif des réponses aux questions 8 à 10, les estimations de ces questions ont été regroupées en une seule question pour l'estimation. 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 moyennes ou de proportions n'ont été produits.

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 B.

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 biaisé³. 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'estimateur⁴.

³ William G. Cochran, *Sampling techniques*, Harvard University, Wiley & sons, 1977 p. 160-161

⁴ William G. Cochran, *Sampling techniques*, Harvard University, Wiley & sons, 1977 p. 162

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

Dans le cadre de la présente enquête, les estimations pour le domaine des établissements provenant des regroupements des secteurs de fabrication, de services ou de fabrication de haute technologie affichent une très bonne précision. En effet, on retrouve beaucoup de cotes A, B ou C pour les estimations de ces domaines. Dans le cas de la fabrication de haute technologie, cette performance peut s'expliquer par le fait qu'une grande proportion d'établissements a été échantillonnée. Des cotes A, B ou C sont aussi très fréquentes dans les autres secteurs de l'enquête, mais il est normal que la précision dans les secteurs désagrégés des services ou de la fabrication soit inférieure à celle de l'estimation provenant du regroupement de ces secteurs. Le plan de sondage a été passablement modifié pour l'enquête de 2005. Ainsi, à la demande du CETECH, une précision accrue pour les totaux a été visée. Le recensement des plus gros établissements a également contribué à diminuer la variabilité. Ainsi, nous observons les résultats de ces changements puisque les marges d'erreurs des totaux sont généralement inférieures (parfois jusqu'à trois fois inférieures) à celles des années passées et plusieurs cotes se sont améliorées. Ceci contribue à obtenir de bonnes précisions même dans les totaux des secteurs plus fins. Il faut

cependant souligner que la taille d'échantillon du secteur de la construction ayant diminué quelque peu, les marges d'erreurs sont légèrement supérieures cette année à ce qu'elles étaient l'an passé.

Cependant, pour les catégories professionnelles des professionnels et des techniciens des sciences naturelles et appliquées, les résultats obtenus cette année affichent une moins bonne précision que ceux de l'an passé. Une très grande variabilité dans les réponses obtenues et la présence dans l'échantillon d'établissements possédant beaucoup de ces emplois semblent les seules causes de cette perte de précision. 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 regroupements de secteurs (fabrication, services, autres fabrication, autres services, etc.) 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 4 et 5 puisque l'on retrouve peu d'établissements qui ont des postes vacants. Ainsi, pour ces questions, il est 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.

Pour la question 5b, quelques établissements ayant déclaré des postes vacants à la question 4b ne savaient pas combien l'étaient depuis au moins quatre mois. Lorsque l'on calcule l'estimation associée à cette question, 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 très faible pour la question 5b car cette situation ne se produit que pour de rares cas. Ce problème était davantage présent lors de l'enquête de 2004. Toutefois, ce phénomène est très présent pour la question 7. Les problèmes liés à l'interprétation de cette question problématique sont discutés à la section 8.2.

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.

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 obtenues plus rarement pour ces régions que pour l'ensemble du Québec. 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.

En raison des modifications dans l'analyse de non-réponse discutées à la section 7.2, on remarque une hausse du nombre d'employés dans les régions centrales et une baisse dans la région de Montréal. Toutefois, ceci n'est certes pas entièrement attribuable à un véritable mouvement sur le terrain. Le changement de méthodologie affecte donc grandement la comparabilité des données avec l'année 2004 au point où ces comparaisons ne devraient pas être faites. Toutefois, puisque la méthodologie sera conservée pour l'année 2006 de l'enquête, les données entre 2005 et 2006 pourront quant à elle être comparées.

Enfin, les estimations par taille d'établissement étant produites seulement pour des secteurs agrégés, la précision des estimations faites sur ces domaines demeure satisfaisante.

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'intervieweur 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 (84,9 %) à l'enquête a été obtenu. De plus, les taux de réponse par strate (voir **tableau 3**) sont aussi satisfaisants puisque, à l'exception de trois strates sur 47, 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, 157 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 2005, 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. La question 7 portant sur le nombre d'employés ayant reçu de la formation fait toutefois exception. En effet, plusieurs répondants n'ont fourni que le nombre de formations dispensées à leurs employés et non le nombre d'employés les ayant suivies. Ils ont donné des chiffres bien supérieurs à ce qu'ils auraient dû être puisqu'un employé peut avoir reçu plusieurs formations. Dans certains cas, le nombre d'employés ayant reçu de la formation était même supérieur au nombre d'employés de l'établissement au cours de l'année 2005. Ce phénomène amène donc une surestimation importante des estimations de la question 7. Par ailleurs, plusieurs autres établissements disaient avoir des employés ayant reçu de la formation à la question 7.1, mais n'étaient pas en mesure de donner de chiffres à la section 7.2. Dans ce cas, aucune formation n'est comptabilisée dans les totaux pour ces établissements et les estimations sont sous-estimées. Ainsi, ces deux phénomènes sont le reflet d'un problème sérieux avec cette question et nous incite à regarder les résultats de cette question avec énormément de prudence.

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'intervieweur 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, 3, 4, 5 et 7), puis, un second type de validation permet de vérifier la cohérence entre les questions. Une telle

validation apparaît notamment à la question 5 afin que les réponses soient cohérentes avec celles de la question 4. 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 autre forme de validation historique a été instaurée depuis l'année dernière. 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 2005, auquel on ajoute les départs en 2005 et on soustrait les embauches en 2005, avec le nombre d'employés au 31 décembre 2004. 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, 148 établissements ont ainsi été rappelés. Parmi ceux qui ont été rejoints, les données de cette année ont été corrigées pour 44 établissements alors que 65 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) de même que pour les secteurs de conception de systèmes informatiques, d'architecture, génie et services connexes ainsi que des services de recherche et de développement scientifiques.

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.

Finalement, des estimations par taille d'établissement sont produites pour une première fois dans le cadre de l'EREQ. Une analyse des problèmes de confidentialité potentiels liés à une telle production a permis de déterminer a priori quelles estimations pouvaient être produites aux niveaux provincial et régional. Ainsi, au niveau provincial, la production par taille d'établissement ne s'effectue que pour les secteurs :

- Construction;
- Fabrication;
- Services;
- Ensemble des secteurs regroupés.

Puis, au niveau régional, la production d'estimations ne se fait que pour l'ensemble des secteurs d'activité regroupés et pour l'ensemble des professions regroupées.

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 permuter 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 supprimés. 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) La méthode pour le calcul des proportions d'établissements

Puisque le fichier de travail contient 6 943 observations provenant des 6 773 établissements échantillonnés, calculer une simple proportion ne serait donc plus une proportion d'établissements, mais une proportion d'observations. Des modifications au processus d'estimation avaient donc dû être apportées lors de l'introduction de la dimension régionale à l'enquête. Ainsi, pour calculer une proportion d'établissements, 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 4 958 enregistrements répondants du fichier, on retrouve le nombre d'établissements répondants de l'échantillon, à savoir 4 788. 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 11, la proportion d'établissements répondants « Moins de 50 % de son temps » peut être calculée à l'aide des variables suivantes :

$$x = \begin{cases} 1 & \text{si Moins de 50\% de son temps} \\ 0 & \text{sinon} \end{cases}$$

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

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.