

Guide de l'utilisateur de microdonnées

Enquête auprès des directeurs d'école

2004-2005

Décembre 2006

Table des matières

1.0 Administration	3
2.0 Loi autorisant la collecte	3
3.0 Contexte	3
4.0 Objectifs	4
5.0 Contenu	4
6.0 Utilisations	5
7.0 Collecte des données	5
8.0 Traitement des données	6
8.1 Saisie des données	6
8.2 Vérification des données et détection des valeurs aberrantes	6
9.0 Techniques d'enquête	7
9.1 Population cible	7
9.1.1 Base de sondage	7
9.2 Plan d'échantillonnage	8
10.0 Qualité des données	10
10.1 Taux de réponse	11
10.2 Erreurs d'échantillonnage	12
10.3 Erreurs non dues à l'échantillonnage	13
10.4 Champ d'observation	13
10.5 Saisie des données	14
10.6 Vérification des données et détection des valeurs aberrantes	14
11.0 Pondération	15
11.1 Pondération de base	15
11.2 Correction de la non-réponse	15
11.3 Étalonnage	16
11.4 Pondération finale	16
12.0 Lignes directrices pour l'analyse, la diffusion et les totalisations	16
12.1 Lignes directrices pour l'arrondissement	17
12.2 Lignes directrices pour la pondération	17
12.3 Estimations nominales ou quantitatives	19
12.3.1 Estimations nominales	19
12.3.2 Estimations quantitatives	20
12.3.3 Totalisation d'estimations nominales	20
12.3.4 Totalisation d'estimations quantitatives	21
12.4 Lignes directrices pour la diffusion du coefficient de variation	23
13.0 Calcul de la variance	25
13.1 Importance de la variance	25
13.2 Calcul de la variance et du CV d'une estimation	25
13.3 Calcul des limites de confiance à l'aide du coefficient de variation	33
13.4 Tests d'hypothèses	34
13.5 Coefficients de variation d'estimations quantitatives	35
Annexe A : Questionnaire	35
Annexe B : Caractéristiques d'une école	35

1.0 Administration

L'Enquête auprès des directeurs d'école (EDE) a été menée conjointement par Statistique Canada et une équipe de chercheurs appartenant aux facultés d'éducation de différentes universités canadiennes (Université de Montréal, Université de Toronto, Université Simon Fraser et Université de Sherbrooke). L'enquête s'inscrit dans un projet de recherche parrainé par le Conseil de recherches en sciences humaines du Canada (CRSH). Le CRSH est un organisme fédéral autonome qui appuie la recherche et la formation avancée en milieu universitaire dans les sciences humaines et qui en fait la promotion.

Pour toute question au sujet de l'ensemble des données ou de son utilisation, veuillez communiquer avec :

Statistique Canada
Services à la clientèle
Centre de la statistique de l'éducation
Téléphone : (613) 951-7608 ou composez sans frais 1 800 307-3382
Télécopieur : (613) 951-9040
Adresse électronique : educationstats@statcan.ca

2.0 Loi autorisant la collecte

L'EDE a été menée en vertu de la *Loi sur la statistique*, chapitre S-19, Lois révisées du Canada, 1985. Les plans de collecte de l'enquête sont conformes aux exigences de la circulaire du Conseil du Trésor n° 1986-19, **Collecte de renseignements et recherche sur l'opinion publique fédérales**, et sont enregistrés sous le matricule de collecte n° STC/ECT-165-75359.

3.0 Contexte

L'EDE, première enquête du genre au Canada, a été menée en 2004-2005 pour recueillir des données sur l'incidence des différents changements observés en éducation, tel que des modifications des programmes, des réductions budgétaires, des nouvelles directives d'orientation sur l'enseignement et le travail des directeurs d'écoles primaires et secondaires du Canada. Conçue à l'initiative d'un groupe de chercheurs universitaires, l'enquête a été menée conjointement par Statistique Canada et une équipe de chercheurs appartenant aux facultés d'éducation de différentes universités canadiennes et s'inscrit dans un projet de recherche parrainé par le Conseil de recherches en sciences humaines du Canada (CRSH), intitulé « Évolution actuelle du personnel de l'enseignement primaire et secondaire au Canada ». Pour en savoir plus, veuillez visiter le site Web du projet :

www.teachcan.ca

4.0 Objectifs

L'objectif principal de cette enquête est d'évaluer l'incidence des différents changements observés en éducation, tel que des modifications des programmes, des réductions budgétaires, des nouvelles directives d'orientation sur l'enseignement et le travail des directeurs d'école du Canada. Les données d'enquête comprennent des renseignements sur les directeurs d'école, leurs conditions de travail et leurs pratiques professionnelles et sur les transformations qui ont eu une incidence sur leur formation professionnelle, leurs compétences, ainsi que sur leur travail et leurs rapports quotidiens avec les élèves et les autres intervenants en éducation.

Ces données seront utiles aux chercheurs, aux décideurs, aux ministères provinciaux et territoriaux de l'éducation, aux conseils et districts scolaires, aux directeurs d'école et aux enseignants et contribueront à définir les politiques à venir en matière d'éducation.

5.0 Contenu

La population cible de l'Enquête auprès des directeurs d'école comprend tous les directeurs d'écoles primaires et secondaires du Canada qui occupaient leur poste au début de l'année scolaire 2004-2005, à l'exception des écoles d'éducation permanente ou de jour pour adultes, des écoles professionnelles ou de métiers, des écoles de langue et d'éducation culturelle, des écoles à domicile, des centres d'éducation communautaire, des centres de services sociaux, des centres d'enseignement à distance, des écoles virtuelles ainsi que des écoles des communautés des Premières nations. Les écoles de toutes les provinces et de tous les territoires sont incluses.

Un échantillon de 4 800 écoles a été invité à participer à cette enquête. La participation était volontaire. Les questionnaires ont été envoyés par la poste aux directeurs d'école à la fin d'octobre 2004.

Le questionnaire est reproduit à l'annexe A.

Le questionnaire était divisé en six sections :

- 1) Renseignements socio-démographiques et caractéristiques de l'école** (questions concernant les caractéristiques sociales des directeurs d'école : âge, niveau d'études, nombre d'années d'expérience, etc.; questions concernant le personnel et l'effectif de l'école : origine sociale des élèves, pourcentage d'élèves qui décrochent, origine ethnique du personnel enseignant, etc.);
- 2) Perception du changement et de ses répercussions** (questions concernant la façon dont les directeurs d'école perçoivent les changements survenus à l'école au cours des dix dernières années et les répercussions de ces changements sur leur travail et sur le fonctionnement de l'école);

3) **Tâches et responsabilités** (questions visant à décrire le travail des directeurs d'école : non seulement le travail qu'ils accomplissent actuellement, mais celui qu'ils souhaiteraient accomplir idéalement);

4) **Rapports sociaux dans les écoles** (questions visant à décrire les rapports qu'entretiennent les divers intervenants au sein de l'école);

5) **L'insertion et le développement professionnel** (questions portant sur le recrutement, l'insertion et le développement professionnel des nouveaux enseignants et des nouveaux directeurs d'école. Par exemple, on demande aux directeurs d'école de quelle façon les nouveaux enseignants sont accueillis à l'école et de quelle façon ils ont bénéficié de cet accueil);

6) **Projets et valeurs éducatives** (questions visant à connaître l'orientation pédagogique de l'école du point de vue des objectifs pédagogiques qui sont mis en valeur).

6.0 Utilisations

Le fichier de microdonnées a été mis à la disposition des chercheurs des facultés d'éducation de l'Université de Montréal, de l'Université de Toronto, de l'Université Simon Fraser et de l'Université de Sherbrooke. L'information servira à évaluer l'incidence des différents changements observés en éducation, tel que des modifications de programme, des réductions budgétaires, des nouvelles directives d'orientation sur l'enseignement et le travail des directeurs d'école du Canada. Statistique Canada envisage de mettre les données à la disposition du public par le biais des centres de données de recherche (CDR).

L'enquête donnera lieu à la publication d'études analytiques liées à l'incidence des différents changements observés en éducation et fournira des renseignements qui contribueront à définir les politiques à venir en matière d'éducation.

7.0 Collecte des données

La collecte des données a eu lieu d'octobre 2003 à février 2004. Les questionnaires ont été envoyés directement par la poste aux directeurs d'école. Ces derniers ont fourni les données dont ils disposaient et exprimé leur opinion sur les conditions d'enseignement. On leur a demandé de remplir le questionnaire et de le retourner par la poste dans l'enveloppe prévue à cette fin.

Même si la participation à l'enquête était volontaire, on a envoyé aux répondants une télécopie de rappel, suivie d'appels téléphoniques pour les inciter à participer. Lorsque l'appel de l'intervieweur n'était pas opportun, on prenait rendez-vous pour rappeler à un autre moment.

Si les répondants n'avaient absolument pas le temps de remplir le questionnaire, on leur offrait alors la possibilité de répondre au plus grand nombre possible de sections du questionnaire, selon un classement par ordre d'importance.

8.0 Traitement des données

Ce chapitre présente un bref aperçu des étapes du traitement qui ont mené à la production du fichier de microdonnées.

8.1 Saisie des données

Les réponses aux questions de l'enquête ont été saisies par la reconnaissance intelligente de caractères (RIC). La technologie de la RIC allie l'entrée de données automatisée (fondée sur la reconnaissance optique de caractères, de marques et d'images) à une saisie manuelle supplémentaire sur clavier par des préposés qui « saisissent de l'image » une partie des informations de l'enquête au moyen d'un affichage frontal.

Pour garantir la qualité des données saisies par la RIC, toutes les zones à remplir faisaient l'objet d'une double saisie au clavier pour en assurer l'exactitude, puis on recourait à une procédure de contrôle qualitatif qui consistait, pour chaque lot de questionnaires saisis, à prélever 20 % des questionnaires pour en comparer l'image aux données réelles.

Des vérifications d'intervalle ont été programmées pour la saisie des données. Si les renseignements entrés se situaient à l'extérieur de l'intervalle (trop grands ou trop petits) des valeurs attendues, ou s'ils étaient incohérents, les données étaient alors vérifiées et modifiées, ou refusées.

8.2 Vérification des données et détection des valeurs aberrantes

Parmi les étapes du traitement, la vérification des données et la détection des valeurs aberrantes consistent à cerner les valeurs erronées ou incohérentes dans les données d'enquête et à les modifier.

Le premier type d'erreur traité consistait à vérifier les enregistrements de l'enquête selon des règles préétablies afin de déceler les incohérences logiques dans les données d'enquête. Dans ces cas, on a mis au point une stratégie en vue d'éliminer un minimum de données, en établissant des priorités de manière à supprimer les valeurs le plus susceptibles d'être erronées. Plusieurs questions comportaient des parties multiples dont la somme des réponses aurait dû donner un total indiqué dans une autre question ou encore un total de 100 % dans les questions où l'on demandait une répartition proportionnelle. On a vérifié ces questions en additionnant simplement les réponses. Dans d'autres cas, il était possible de déduire qu'une réponse devait être inférieure à une quantité indiquée pour une question connexe. Dans ces cas, on a identifié les erreurs chaque fois que la deuxième valeur était supérieure à la première. Par exemple, le nombre d'élèves d'une école donnée qui ont accès à des services spécialisés ne peut excéder le nombre total d'élèves de cette école. Dans l'exemple mentionné, on a découvert que certains directeurs d'école avaient déclaré par erreur des pourcentages et l'on a imputé la valeur pertinente en fonction de la réponse donnée; si le nombre total d'élèves de l'école est de deux

cents et que le directeur, au lieu de donner un nombre, a déclaré par erreur que 50 % des élèves avaient accès à des services spécialisés, on a imputé la valeur de $100 = 50 \% * 200$.

La deuxième composante du processus de vérification visait le traitement des réponses multiples aux questions du type de l'échelle de Likert (questions 1 à 3 de la section 4, par exemple). À ces questions, certains répondants cochaient plus d'une case. Pour chacun de ces enregistrements à réponses multiples, on a retenu la première réponse ou la réponse la plus à gauche.

Le troisième type d'erreur traité consistait à vérifier les enregistrements présentant des valeurs inusitées. On a détecté les erreurs selon des règles préétablies, comme des vérifications de ratios, et par une vérification des plus grandes et des plus petites valeurs. À titre d'exemple d'une règle préétablie, on a identifié les écoles ayant indiqué que plus de la moitié de leur population étudiante était arrivée au Canada depuis moins d'un an. La vérification était fondée sur le type d'école et sur les écoles semblables des environs.

9.0 Techniques d'enquête

Ce chapitre définit la population d'intérêt, les critères de sélection de cette population et la méthode de communication avec elle, ainsi que la méthode d'enquête employée.

9.1 Population cible

La population cible de l'Enquête auprès des directeurs d'école (EDE) comprend tous les directeurs d'écoles primaires et secondaires du Canada qui occupaient leur poste au début de l'année scolaire 2004-2005, à l'exception des écoles d'éducation permanente ou de jour pour adultes, des écoles professionnelles ou de métiers, des écoles de langue et d'éducation culturelle, des écoles à domicile, des centres d'éducation communautaire, des centres de services sociaux, des centres d'enseignement à distance, des écoles virtuelles ainsi que des écoles des communautés des Premières nations. Les écoles de toutes les provinces et de tous les territoires sont incluses.

9.1.1 Base de sondage

On peut avoir accès à la population grâce à la base de sondage. Pour l'EDE de 2005, on a utilisé comme base de sondage une liste d'écoles, car on ne disposait pas de liste de directeurs d'école. La base de sondage utilisée est le fichier des établissements pour 2002, une base de données administratives de toutes les écoles primaires et secondaires du Canada mise à jour par le Centre de la statistique de l'éducation (CSE) de Statistique Canada. Le fichier des établissements renferme un ensemble de renseignements sur chaque école, notamment la personne-ressource, le niveau d'année minimal et maximal, le financement de l'école et le nombre d'élèves qui fréquentent l'école. Le CSE rassemble ces informations à partir des sources les plus à jour qui déclarent des données annuelles sur les écoles primaires et secondaires du Canada, afin d'actualiser le fichier des établissements de l'année précédente. Des demandes sont envoyées aux ministères provinciaux et (ou) aux directeurs d'école pour

mettre à jour l'information de chaque école dans le fichier. Le fichier des établissements utilisé pour créer la base de sondage de l'EDE contenait toutes les mises à jour reçues jusqu'à l'été 2004 pour les écoles déjà comprises dans la base de sondage. On a également modifié la base de sondage de l'EDE pour éliminer les écoles hors du champ de l'enquête qui ont été repérées au moment de l'Enquête sur les technologies de l'information et des communications dans les écoles (ETICE) de 2003-2004, qui utilisait comme base de sondage une version antérieure du même fichier des établissements.

9.2 Plan d'échantillonnage

Compte tenu des objectifs, des contraintes et du budget de cette enquête, on a décidé de sélectionner un échantillon d'écoles au lieu d'utiliser un recensement des écoles. Afin d'obtenir des estimations fiables au sein de certaines sous-populations d'intérêt pour l'enquête, on a stratifié la liste des écoles en fonction de la région et du niveau d'année des écoles, soit les six régions et les trois niveaux d'année suivants :

Six régions

- Provinces de l'Atlantique : Terre-Neuve-et-Labrador, Nouveau-Brunswick, Nouvelle-Écosse et Île-du-Prince-Édouard
- Québec
- Ontario
- Provinces des Prairies : Manitoba, Saskatchewan et Alberta
- Colombie-Britannique
- Territoires : Yukon, Territoires du Nord-Ouest et Nunavut

Trois niveaux d'année

- Primaire
- Secondaire
- Mixte (primaire et secondaire)

On a ainsi créé 16 strates : trois niveaux d'année par région pour toutes les régions, sauf les territoires, où l'on a considéré l'ensemble des écoles.

Tableau 9.1
Stratification

1- Atlantique-Primaire
2- Atlantique-Secondaire
3- Atlantique-Mixte
4- Québec-Primaire
5- Québec-Secondaire
6- Québec-Mixte
7- Ontario-Primaire
8- Ontario-Secondaire
9- Ontario-Mixte
10- Prairies-Primaire

11- Prairies–Secondaire
12- Prairies–Mixte
13- C.-B.–Primaire
14- C.-B.–Secondaire
15- C.-B.–Mixte
16- Territoires

Pour cette enquête, on a ciblé un échantillon de 2 000 répondants. En se fondant sur le taux de réponse régional de l'ETICE, on a déterminé qu'un échantillon de 4 800 écoles serait sans doute suffisant pour produire les 2 000 répondants recherchés.

On a réparti l'échantillon de 4 800 écoles entre les 16 strates afin d'obtenir, dans le plus grand nombre de strates possible, des coefficients de variation (CV) inférieurs ou égaux à 16 % pour des proportions parfois aussi faibles que 10 %. Étant donné le petit nombre d'écoles dans les territoires, on a mené un recensement des écoles dans cette strate. On a réparti les écoles restantes entre les 15 autres strates selon une approche en deux étapes : une première étape de répartition, proportionnelle à la racine carrée du nombre d'écoles de chaque région, suivie d'une deuxième étape, proportionnelle au nombre d'écoles de chaque niveau d'année à l'intérieur de chaque région. Par rapport aux autres approches possibles, cette répartition a donné de meilleurs CV globaux à l'échelle régionale et nationale sans avoir d'incidence significative sur la fiabilité des estimations concernant les petites strates. Les tableaux 9.2 et 9.3 montrent la taille de la population et l'échantillon répartis par région et par niveau d'année.

Tableau 9.2
Nombre d'écoles par région et par niveau d'année

Population	Primaire	Secondaire	Mixte	Total
Atlantique	689	317	202	1 208
Québec	2 216	604	139	2 959
Ontario	4 182	1 134	231	5 547
Prairies	1 525	745	1 165	3 435
Colombie-Britannique	1 305	494	175	1 974
Territoires	47	20	54	121
Total	9 964	3 314	1 966	15 244

Tableau 9.3
Taille de l'échantillon par région et par niveau d'année

Échantillon	Primaire	Secondaire	Mixte	Total
Atlantique	349	161	102	612
Québec	716	195	45	956
Ontario	981	266	54	1 301
Prairies	455	223	348	1 026
Colombie-Britannique	518	196	70	784
Territoires	47	20	54	121
Total	3 066	1 061	673	4 800

Sélection de l'échantillon

À l'intérieur de chacune des 16 strates créées par région et par niveau d'année, on a trié les écoles en fonction des variables auxiliaires clés, dans l'ordre suivant : langue d'enseignement, type de financement (public ou privé), taille (nombre total d'élèves) et emplacement (urbain ou rural). On a ensuite obtenu l'échantillon au moyen d'un échantillonnage systématique des écoles à l'intérieur de chaque strate. Cette méthode d'échantillonnage garantit que les écoles retenues sont représentatives des variables principales explicites (région et niveau d'année), ainsi que des variables clés implicites utilisées pour trier la base de sondage avant la sélection.

L'utilisateur doit prendre note que les écoles inscrites dans la base de sondage étaient les unités échantillonnées, mais que les directeurs d'école étaient les unités répondantes. Si un directeur d'école était responsable de plusieurs écoles sélectionnées dans l'échantillon, il devait répondre pour chacune des écoles retenues puisque certaines questions concernaient spécifiquement l'école.

10.0 Qualité des données

Ce chapitre permet à l'utilisateur de prendre connaissance des divers éléments qui influencent la qualité des données de l'enquête. Dans une enquête par sondage ordinaire, on distingue deux principaux types d'erreur : les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage. On appelle erreur d'échantillonnage la différence entre une estimation dérivée d'un échantillon et l'estimation qu'on aurait obtenue à partir d'un recensement qui aurait utilisé les mêmes procédés de collecte. On appelle erreurs non dues à l'échantillonnage tous les autres types d'erreur, y compris les problèmes liés au champ d'observation de l'enquête, la non-réponse et les erreurs de traitement, qui sont examinées dans les sections ci-après.

10.1 Taux de réponse

Les taux de réponse d'une enquête mesurent l'efficacité du processus de collecte et sont aussi de bons indicateurs de la qualité des estimations produites. En raison des contraintes de temps liées au travail des directeurs d'école et du nombre d'enquêtes auxquelles ils ont à répondre, il est difficile d'obtenir d'eux des taux de réponse élevés. On a donc mis au point plusieurs méthodes pour encourager la réponse. On a avisé les conseils scolaires à l'avance de l'enquête et, 21 jours après l'envoi des questionnaires, on a envoyé un rappel par télécopieur aux écoles qui n'avaient pas encore répondu. Dix jours après l'envoi de ce rappel, on a effectué un suivi téléphonique en offrant aux répondants la possibilité de transmettre leurs réponses par télécopieur, plutôt que par la poste, afin d'accroître le taux de réponse. On a obtenu un taux de résolution final de 56 %, un cas étant résolu lorsqu'on avait communiqué avec le répondant éventuel ou déterminé que l'école se trouvait hors du champ de l'enquête. On entend par répondant une personne ayant rempli au moins la section 1 de l'enquête. Sur les 4 658 directeurs d'école de l'échantillon qui se trouvaient dans le champ de l'enquête, 2 230 ont répondu, ce qui donne un taux de réponse de 47,9 %. Sur les 2 230 directeurs d'école ayant répondu, 2 144 ont consenti à la divulgation de leurs renseignements aux universités qui parrainent cette étude, soit un taux de consentement de 96,1 %.

Tableau 10.1
Nombre de cas résolus

Échantillon	Cas résolus			Cas non résolus (sans réponse)	Taux de résolution
	Hors du champ de l'enquête	Refus	Réponses		
4 800	142	329	2 230	2 099	56 %

Tableau 10.2
Taux de réponse par région et par niveau d'année

Région	Niveau d'année	Écoles dans le champ de l'enquête	Répondants		Consentants	
			Nombre	Taux	Nombre	Taux
			%		%	
Canada	Total	4 658	2 230	47,9	2 144	96,1
	Primaire	2 978	1 408	47,3	1 352	96,0
	Mixte	649	297	45,8	283	95,3
	Secondaire	1 031	525	50,9	509	97,0
Atlantique	Total	596	345	57,9	330	95,7
	Primaire	341	199	58,4	192	96,5
	Mixte	99	50	50,5	48	96,0
	Secondaire	156	96	61,5	90	93,8
Québec	Total	939	437	46,5	421	96,3
	Primaire	705	320	45,4	306	95,6
	Mixte	44	19	43,2	19	100,0
	Secondaire	190	98	51,6	96	98,0
Ontario	Total	1 260	575	45,6	546	95,0
	Primaire	950	426	44,8	406	95,3
	Mixte	49	22	44,9	19	86,4
	Secondaire	261	127	48,7	121	95,3
Prairies	Total	998	501	50,2	482	96,2
	Primaire	446	229	51,3	220	96,1
	Mixte	336	152	45,2	144	94,7
	Secondaire	216	120	55,6	118	98,3
C.-B.	Total	749	315	42,1	309	98,1
	Primaire	493	211	42,8	205	97,2
	Mixte	67	28	41,8	28	100,0
	Secondaire	189	76	40,2	76	100,0
Territoires	Total	116	57	49,1	56	98,2
	Primaire	43	23	53,5	23	100,0
	Mixte	54	26	48,1	25	96,2
	Secondaire	19	8	42,1	8	100,0

On appelle taux de réponse d'item le taux de réponse à chaque question. Les taux de réponse d'item ont été très élevés : ils étaient tous supérieurs à 96 %, sauf ceux de deux questions, qui se situaient entre 85 % et 90 %.

10.2 Erreurs d'échantillonnage

Les estimations dérivées de cette enquête sont fondées sur un échantillon d'écoles. On appelle erreur d'échantillonnage de l'estimation la différence entre les estimations obtenues à partir de l'échantillon et les résultats d'un recensement complet mené dans des conditions semblables. Comme les estimations dérivées d'une enquête par sondage sont inévitablement sujettes à l'erreur d'échantillonnage, une saine pratique de la statistique exige que les chercheurs

renseignent les utilisateurs sur l'ampleur de cette erreur. Cette section de la documentation présente les mesures de l'erreur d'échantillonnage que Statistique Canada utilise couramment et dont il conseille vivement l'utilisation aux personnes qui produisent des estimations à partir de ce fichier de microdonnées.

On mesure l'ampleur éventuelle des erreurs d'échantillonnage en fonction de l'erreur-type des estimations dérivées des résultats de l'enquête. Toutefois, en raison de la grande diversité des estimations qu'on peut produire à partir d'une enquête, on exprime habituellement l'erreur-type d'une estimation en fonction de l'estimation à laquelle elle se rapporte. On obtient le coefficient de variation (CV) d'une estimation en exprimant l'erreur-type de l'estimation en pourcentage de l'estimation proprement dite. Pour en savoir plus sur les CV, veuillez consulter la section 12.4 et le chapitre 13.

10.3 Erreurs non dues à l'échantillonnage

Il y a de nombreuses sources d'erreurs non dues à l'échantillonnage et elles peuvent se produire à presque toutes les étapes d'une enquête. Par exemple : des employés responsables de la collecte comprennent mal les instructions; des répondants ne comprennent pas des questions; ou des réponses sont mal inscrites sur le questionnaire. Des erreurs peuvent se produire à tout moment lors de la manipulation des données, y compris lors du dépouillement et de la totalisation des données.

Pour l'EDE, on a adopté des mesures d'assurance de la qualité afin de réduire le plus possible la présence d'erreurs non dues à l'échantillonnage. Des consultations ont eu lieu à l'automne 2003 à Moncton, à Montréal et à Toronto, et 30 directeurs y étaient pour évaluer, individuellement et en groupes de discussion, le questionnaire. La composition de l'échantillon des groupes de discussion a tenu compte des caractéristiques suivantes : la langue de l'école, le niveau de l'école, les écoles à forte concentration de minorités visibles, le financement accordé à l'école, le sexe et le nombre d'années d'expérience du directeur. De plus, environ 200 directeurs de différentes localités au Canada ont été invités à participer à une enquête pilote au printemps 2004. Des changements ont été apportés au questionnaire après les rencontres des groupes de discussion et l'essai pilote. Ces changements visaient à réduire le fardeau des répondants et à améliorer la qualité des données.

Pour plus de renseignements sur les procédures de collecte et de traitement, veuillez consulter les chapitres 7.0 et 8.0. Quelques-unes de ces mesures indiquent l'étendue des erreurs non dues à l'échantillonnage associées à l'enquête et elles sont décrites ci-après.

10.4 Champ d'observation

On a évalué la qualité de la base de sondage en examinant le sous-dénombrement et le surdénombrement, de même que les enregistrements en double. Des difficultés sont survenues lors de l'évaluation du sous-dénombrement de la base de sondage, car on a découvert que la plupart des autres listes auxquelles on pouvait comparer le fichier des établissements étaient soit étroitement liées aux fichiers administratifs utilisés pour créer la base de sondage de

l'EDE, soit apparentées à une version antérieure de ces fichiers, soit dépourvues d'un indicateur commun. La base de sondage était à jour dans le cas des écoles ouvertes au début de l'année scolaire 2003-2004, l'information sur ces écoles ayant été mise à jour pendant l'été 2004, mais il y avait un degré indéterminé de sous-dénombrement des écoles qui pouvaient avoir ouvert leurs portes pendant la période comprise entre le début de l'année scolaire 2003-2004 et celui de l'année scolaire 2004-2005.

10.5 Saisie des données

Les réponses fournies sur le questionnaire par les directeurs d'école ont été saisies au moyen de l'imagerie électronique. Les seules exceptions étaient les questions auxquelles les directeurs pouvaient répondre sous forme de texte ou certains questionnaires, dont des questionnaires télécopiés, ayant causé des problèmes au moment de la saisie électronique des données; dans ces cas, on a saisi les données au clavier sur micro-ordinateur. On a utilisé un logiciel de reconnaissance de textes pour saisir les données et toutes les réponses saisies, manuellement ou électroniquement, ont fait l'objet d'une double vérification.

10.6 Vérification des données et détection des valeurs aberrantes

Les processus de vérification des données et de détection des valeurs aberrantes sont décrits au chapitre 8.0.

Une grande partie de la vérification a porté sur la section 1 du questionnaire, qui présentait le plus d'aspects problématiques. Les questions 10 à 13 et 15 ont fait l'objet d'une vérification approfondie. Étant donné la piètre qualité des données concernant le personnel à temps partiel aux questions 9 et 14 de la section 1, les données de ces questions n'ont pas été corrigées et ne devraient pas être utilisées à des fins analytiques. Les données d'environ 80 % des enregistrements de la section 1 du questionnaire ont fait l'objet d'au moins une correction.

La dernière question concernant la divulgation des données a également fait l'objet d'un examen approfondi. On a décelé un certain nombre d'erreurs de saisie, notamment des cas dans lesquels il y avait une coche à l'extérieur de la boîte, une coche dans la boîte mais non saisie, ou encore aucune coche alors que le répondant avait signé. Ces enregistrements ont été corrigés manuellement après avoir été examinés cas par cas.

11.0 Pondération

11.1 Pondération de base

Dans un échantillon probabiliste, le plan d'échantillonnage détermine les pondérations initiales à utiliser pour produire des estimations non biaisées de la population. Chaque unité doit être pondérée par l'inverse de sa probabilité de sélection. Dans le cas d'un échantillon aléatoire simple à 2 %, cette probabilité serait de 0,02 pour chaque unité; les pondérations initiales seraient donc de $1/0,02=50$. Pour cette enquête, la probabilité de sélection variait par strate (région par niveau d'année) et les pondérations initiales ont été calculées comme suit :

Pondération de base $W(1)$:

$$W(1) = \frac{\text{Nombre d'écoles de la strate dans la base de sondage}}{\text{Nombre d'écoles de la strate dans l'échantillon}}$$

11.2 Correction de la non-réponse

La non-réponse est la principale source d'erreur dans une enquête comme l'EDE. Les erreurs de cette nature n'étant généralement pas aléatoires, il importe de réduire la non-réponse et aussi d'adopter une stratégie de correction appropriée qui tienne compte de la non-réponse systématique pour éviter que les estimations soient fortement biaisées. À l'étape de la collecte des données, on a adopté diverses stratégies (perfectionnement du questionnaire après un essai préliminaire sur le terrain, prise de contact au moyen d'une lettre de présentation, suivi auprès des non-répondants et possibilité de retourner les questionnaires par la poste ou par télécopieur) afin d'améliorer la réponse.

On a examiné et comparé plusieurs méthodes de traitement de la non-réponse. La régression logistique a révélé que les variables les plus étroitement liées à la non-réponse étaient la région, le niveau d'année et la taille. On a comparé différents groupements manuels et différentes grappes probabilistes d'écoles en utilisant des coefficients de variation (CV) pour déterminer lesquels réduiraient le mieux le biais de non-réponse. On a utilisé une mesure diagnostique présentée dans l'article « Mieux comprendre la transformation des poids à l'aide d'une mesure de changement »¹ pour évaluer les différences entre les principales méthodes. Cette mesure diagnostique produit une quantité donnée indiquant la réduction relative du biais de non-réponse produite par un modèle de pondération. On a fini par déterminer qu'un modèle comprenant 54 catégories fondées sur l'intégration des variables significatives *région* (Atlantique, Québec, Ontario, Prairies, Colombie-Britannique et territoires), *niveau d'année* (primaire, secondaire et mixte) et *taille* (petite, moyenne et grande) constituait la meilleure solution pour corriger les facteurs de pondération de manière à compenser la non-réponse. Ce

¹ Dufour, J., Gagnon, F., Morin, Y., Renaud, M. et C.-E. Särndal (2001). Mieux comprendre la transformation des poids à l'aide d'une mesure de changement. *Techniques d'enquête*, vol. 27, n° 1, p. 97 à 108.

modèle offrait en outre l'avantage de ne nécessiter aucun étalonnage supplémentaire, comme nous le montrons dans la prochaine section. La non-réponse des répondants d'une catégorie de correction donnée est donc corrigée comme suit :

Correction de la non-réponse W(2) :

$$W(2) = \frac{\text{Nombre d'unités échantillonnées de cette catégorie dans l'échantillon}}{\text{Nombre d'unités échantillonnées répondantes dans cette catégorie}}$$

11.3 Étalonnage

Les techniques d'estimation par étalonnage sont très utilisées dans les enquêtes sociales. Elles produisent des estimations des totaux pour les variables clés qui correspondent aux chiffres de population connus. Cette approche permet aussi d'améliorer la qualité des estimations de l'enquête s'il y a un rapport entre les variables clés utilisées dans l'étalonnage et les variables utilisées pour l'estimation. Pour cette enquête, il fallait que les totaux correspondent aux chiffres de population par région et par niveau d'année. Comme l'échantillon était stratifié à ces niveaux et que les catégories de non-réponse étaient aussi définies à ces niveaux (voir plus haut), aucun étalonnage supplémentaire n'était nécessaire.

11.4 Pondération finale

En général, le facteur de pondération final est le produit des facteurs qui correspondent aux différentes étapes de sélection de l'échantillon ou aux ajustements comme la pondération initiale, la correction de la non-réponse et l'étalonnage. Dans l'enquête qui nous occupe, il n'y a que la pondération de base et la correction de la non-réponse, car aucun étalonnage n'était nécessaire. La variable pondération finale, appelée « cw » dans le fichier, est donc calculée comme suit :

$$\text{Facteur de pondération final} = W(1) \times W(2)$$

12.0 Lignes directrices pour l'analyse, la diffusion et les totalisations

Ce chapitre donne une vue d'ensemble des lignes directrices que doivent observer les utilisateurs lors de la totalisation, de l'analyse, de la publication ou de la diffusion de données provenant du fichier de microdonnées de l'EDE. Les utilisateurs des microdonnées qui suivront ces lignes directrices devraient pouvoir produire les mêmes statistiques que Statistique Canada.

12.1 Lignes directrices pour l'arrondissement

Afin que les estimations destinées à la publication ou à toute autre forme de diffusion qui sont tirées des fichiers de microdonnées de l'EDE correspondent à celles de Statistique Canada, nous conseillons vivement aux utilisateurs de se conformer aux lignes directrices ci-après en ce qui concerne l'arrondissement de ces estimations.

Les estimations dans le corps d'un tableau statistique devraient être **arrondies à la centaine près** au moyen de la technique d'arrondissement classique. Dans cette technique, si le premier chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier chiffre à supprimer se situe entre 5 et 9, on augmente de 1 la valeur du dernier chiffre à conserver. Par exemple, lorsqu'on cherche à arrondir à la centaine près de façon classique, si les deux derniers chiffres se situent entre 00 et 49, il faut les remplacer par 00 et laisser le chiffre précédent (celui des centaines) tel quel. Si les deux derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on augmente de 1 le chiffre précédent.

Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques devraient être calculés à partir de leurs composantes correspondantes non arrondies, puis arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement classique.

Les moyennes, les proportions, les taux et les pourcentages devraient être calculés à partir d'éléments non arrondis (c.-à-d. les numérateurs et (ou) les dénominateurs), puis **arrondis à une décimale** au moyen de la technique d'arrondissement classique. Lorsqu'on cherche à arrondir à un seul chiffre par cette technique, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, le dernier chiffre à conserver ne change pas. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on augmente de 1 le dernier chiffre à conserver.

Les sommes et les différences d'agrégats (ou de rapports) devraient être calculées à partir de leurs éléments correspondants non arrondis, puis arrondis à leur tour à la centaine près (ou à la décimale près) au moyen de la technique d'arrondissement classique.

Si, en raison de limitations d'ordre technique ou de toute autre nature, on utilise une autre technique d'arrondissement que l'arrondissement classique, de sorte que les estimations à publier ou à diffuser sous une forme quelconque diffèrent des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces divergences dans le ou les documents à publier ou à diffuser.

Des estimations non arrondies ne devraient en aucun cas être publiées ou diffusées par les utilisateurs.

12.2 Lignes directrices pour la pondération

Pour l'EDE, on a sélectionné un échantillon d'écoles au moyen de l'échantillonnage stratifié (pour plus de détails, voir la section 9.2) et on a obtenu un taux de réponse de 48 % à l'étape de la collecte. On doit donc tenir compte de la pondération de base ainsi que de la pondération de la non-réponse lorsqu'on produit des estimations. La pondération de base et la correction de la

non-réponse sont combinées pour former la variable appelée « Pondération finale », qui est mise à la disposition des utilisateurs.

Pour produire des estimations ponctuelles, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer la pondération d'enquête décrite au chapitre 11. Si l'on n'utilise pas la pondération d'enquête, les estimations dérivées des microdonnées ne peuvent être considérées comme représentatives de la population cible et ne correspondront pas à celles produites par Statistique Canada. On peut considérer la pondération attribuée à chaque directeur (école) comme le nombre de directeurs (écoles) dans la base de sondage représentés par ce répondant particulier.

Pour certaines techniques d'analyse (par exemple, la régression linéaire, la régression logistique, l'analyse de la variance ou toute autre analyse qui nécessite une mesure significative), on recommande d'utiliser un facteur de pondération corrigé. La méthode utilisée à cette fin rééchelonne les facteurs de pondération de sorte que le poids moyen soit 1 et que la somme de tous les poids corresponde au nombre de répondants.

Supposons, par exemple, qu'il faille analyser tous les directeurs d'école secondaire. Voici les étapes à suivre pour remettre à l'échelle les facteurs de pondération :

- Sélectionner du fichier tous les répondants qui ont été classés comme directeurs d'école secondaire.
- Calculer la pondération **moyenne** pour ces enregistrements en additionnant les facteurs de pondération finals des directeurs d'école, à partir du fichier de microdonnées pour ces enregistrements, puis les diviser par le nombre de répondants classés comme directeurs d'école secondaire.
- Pour chacun de ces répondants, calculer un facteur de pondération rééchelonné, égal à la pondération finale des directeurs d'école, et le diviser par le facteur de pondération moyen.
- Effectuer l'analyse, pour ces répondants, à l'aide du facteur de pondération **rééchelonné**.

Cette méthode peut être utile pour interpréter les résultats provenant d'un logiciel statistique ordinaire. Bien que de nombreuses méthodes d'analyse faisant partie de progiciels statistiques permettent d'utiliser des pondérations, celles-ci diffèrent, sous l'angle de leur définition ou de leur signification, de celles qui conviennent dans le cadre d'une enquête comme l'EDE. En effet, si les estimations ponctuelles produites au moyen de ces progiciels sont exactes dans bien des cas, les estimations de la variance calculées n'ont à peu près pas de signification. De plus, ces estimations risquent de ne pas concorder avec celles de Statistique Canada, en raison de la manière dont certains progiciels statistiques traitent la zone du facteur de pondération. Le rééchelonnage des facteurs de pondération peut rendre plus significatives les variances dégagées par les logiciels ordinaires. Un des avantages de la correction des facteurs de pondération est qu'elle permet d'éviter la surestimation du niveau de signification, lequel

dépend fortement de la taille de l'échantillon, tout en conservant les mêmes distributions que celles obtenues au moyen du facteur de pondération final.

Lorsqu'on utilise des logiciels statistiques ordinaires pour calculer des estimations assorties à des facteurs de pondération rééchelonnés, on ne tient pas toujours compte de la variabilité des estimations inhérentes aux probabilités de sélection et au mécanisme de non-réponse. C'est pourquoi les estimations de la variance ainsi calculées sous-estiment vraisemblablement la variance véritable. Pour calculer des estimations de la variance plus précises, il faut connaître plus intimement la stratification de l'enquête et le comportement présumé de la population observée en ce qui concerne la non-réponse de même que le procédé de correction correspondant qui a été choisi. Pour plus de renseignements sur le calcul des estimations de la variance pour l'EDE, veuillez consulter le chapitre 13.0.

12.3 Estimations nominales ou quantitatives

L'unité d'analyse dans le fichier de l'EDE est le directeur d'école. Même si l'on peut tirer d'autres estimations de l'EDE, les facteurs de pondération de chaque enregistrement ont été conçus pour fournir des estimations fiables de proportions à partir des chiffres des directeurs d'école (c.-à-d. ceux des écoles). Par exemple, l'estimation de la proportion de directeurs qui déclarent être nés une certaine année est plus fiable que l'estimation de la proportion d'élèves des écoles dont le directeur est né cette année-là.

Avant d'aborder la façon de totaliser et d'analyser les données de l'EDE, il convient de décrire les deux principaux types d'estimation ponctuelle des caractéristiques de la population que le fichier de microdonnées de l'EDE permet de produire. Pour simplifier, on suppose que l'ensemble des répondants appropriés a été utilisé dans les descriptions suivantes. De plus, nous entendrons par **domaine** un groupe de directeurs d'école pour lesquels on produira une estimation (par exemple, l'ensemble des directeurs d'écoles francophones situées dans une province ou un territoire particulier).

12.3.1 Estimations nominales

Les estimations nominales sont des estimations du nombre ou du pourcentage de personnes, dans la population observée, qui possèdent certaines caractéristiques ou qui font partie d'une catégorie définie.

Exemples de questions nominales :

Q : Quelle langue parlez-vous le plus souvent à la maison?

R : Français / Anglais / Autre – précisez

Q : Présentement, y a-t-il à votre école une politique écrite d'évaluation des élèves?

R : Oui / Non

12.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures de tendance centrale de quantités reposant sur une partie ou la totalité des membres de la population observée. Elles comprennent aussi explicitement des estimations de la forme \hat{X}/\hat{Y} , où \hat{X} est une estimation de la quantité totale pour la population observée et \hat{Y} est une estimation du nombre de directeurs d'école dans la population observée qui contribuent à cette quantité totale, ou d'autres estimations plus complexes comme les coefficients de régression.

Voici un exemple d'estimation quantitative : le nombre moyen d'années d'expérience à titre de directeur d'école primaire. Le numérateur est une estimation du nombre total d'années d'expérience de tous les directeurs à titre de directeur d'école primaire et le dénominateur est une estimation du nombre total de directeurs d'école primaire.

Exemples de questions quantitatives :

Q : Quel est le nombre total approximatif d'étudiants inscrits à l'école?

R : Total

Q : Parmi les élèves et le personnel enseignant (temps plein et temps partiel) de votre école, environ combien d'entre eux ont comme langue maternelle une langue autre que le français ou l'anglais?

R : Élèves Personnel enseignant

12.3.3 Totalisation d'estimations nominales

Nombre de directeurs d'école présentant une certaine caractéristique

Lorsque tous les répondants du domaine ont répondu à la question liée à la caractéristique d'intérêt (c.-à-d. qu'il n'y a aucune non-réponse à une question), on peut obtenir l'estimation à partir du fichier de microdonnées en additionnant les facteurs de pondération finals de tous les enregistrements du domaine affichant cette caractéristique.

S'il y a une non-réponse dans le domaine pour la question pertinente, on peut obtenir l'estimation en deux étapes, décrites ci-après. Cette approche convient uniquement lorsque la non-réponse observée à une question est faible (de préférence, moins de 5 %).

Étape 1 : Calculer la proportion des directeurs du domaine ayant répondu présentant cette propriété (voir le procédé ci-après concernant les proportions).

Étape 2 : Multiplier cette proportion par la somme des facteurs de pondération pour tous les directeurs du domaine.

Proportions des directeurs d'école présentant une certaine caractéristique

- a) Additionner les facteurs de pondération finals des directeurs du domaine présentant la caractéristique d'intérêt (\hat{X}).
- b) Additionner les facteurs de pondération finals de tous les directeurs du domaine (\hat{Y}), en excluant ceux qui n'ont pas répondu à la question qui servait à identifier la caractéristique.
- c) Diviser l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

12.3.4 Totalisation d'estimations quantitatives

Estimation des chiffres de population

Les chiffres de population des variables quantitatives désignent des valeurs telles que le nombre total d'élèves ou le nombre total d'enseignants. S'il n'y a pas de non-réponse pour la variable d'intérêt pour tous les répondants (directeurs d'école) du domaine, on peut obtenir des estimations des chiffres de population à partir du fichier de microdonnées comme suit :

- a) Multiplier la valeur de la variable d'intérêt par le facteur de pondération final pour chaque directeur du domaine.
- b) Additionner cette quantité pour tous les directeurs du domaine.

Comme c'était le cas pour les estimations nominales, s'il n'y a pas de non-réponse dans un domaine particulier pour la variable d'intérêt, on doit alors dériver l'estimation en deux étapes, décrites ci-après.

Étape 1 : Obtenir la valeur moyenne de la quantité qui nous intéresse pour les directeurs ayant répondu dans le domaine en utilisant la procédure ci-après pour les ratios.

Étape 2 : Multiplier cette moyenne par le total des facteurs de pondération pour tous les directeurs (ayant répondu ou non) du domaine.

Veillez noter que cette approche en deux étapes ne s'applique que si la non-réponse à la question pertinente dans le domaine qui nous intéresse est faible.

Ratios de la forme \hat{X}/\hat{Y}

Les estimations par ratio représentent plusieurs types d'estimation. Elles comprennent des proportions, des moyennes (par ex., l'âge moyen des directeurs d'école) et des ratios de deux

variables quantitatives (par ex., le ratio du nombre d'élèves divisé par le nombre d'enseignants). On obtient les estimations par ratio comme suit :

a) Multiplier la valeur de la variable d'intérêt pour le numérateur par le facteur de pondération final et additionner cette quantité pour tous les enregistrements du domaine afin d'obtenir le numérateur (\hat{X}), en excluant ceux qui n'ont pas répondu aux variables d'intérêt définissant le numérateur et le dénominateur.

b) Multiplier la valeur de la variable d'intérêt pour le dénominateur par le facteur de pondération final et additionner cette quantité pour tous les enregistrements du domaine afin d'obtenir le dénominateur (\hat{Y}), en excluant ceux qui n'ont pas répondu aux variables d'intérêt définissant le numérateur et le dénominateur.

c) Diviser l'estimation a) par l'estimation b) (\hat{X} / \hat{Y}).

Par exemple, pour estimer l'âge moyen des directeurs d'école primaire :

- Multiplier l'âge de chaque directeur d'école primaire par le facteur de pondération final, puis additionner cette quantité pour tous les directeurs d'école primaire (\hat{X}), en excluant ceux qui n'ont pas indiqué leur année de naissance.
- Additionner les facteurs de pondération finals de tous les directeurs d'école primaire (\hat{Y}), en excluant ceux qui n'ont pas indiqué leur année de naissance.
- Diviser la première estimation par la deuxième estimation (\hat{X} / \hat{Y}).

Un autre exemple est l'estimation du ratio du nombre d'années d'expérience à titre de directeur d'école au nombre total d'années d'expérience à titre d'enseignant, de directeur adjoint, de directeur ou de conseiller pédagogique :

- Multiplier le nombre d'années d'expérience à titre de directeur d'école par le facteur de pondération final, puis additionner cette quantité pour tous les directeurs d'école (\hat{X}), en excluant ceux qui n'ont pas répondu à la question concernant leurs années d'expérience.
- Multiplier le nombre total d'années d'expérience pour l'ensemble de la carrière à titre d'enseignant, de directeur adjoint, de directeur ou de conseiller pédagogique par le facteur de pondération final, puis additionner cette quantité pour tous les directeurs d'école (\hat{Y}), en excluant ceux qui n'ont pas répondu à la question concernant leurs années d'expérience.
- Diviser la première estimation par la deuxième estimation (\hat{X} / \hat{Y}).

12.4 Lignes directrices pour la diffusion du coefficient de variation

Avant de diffuser et (ou) de publier des estimations tirées de l'EDE, les utilisateurs devraient d'abord établir le niveau de qualité de l'estimation. Les niveaux de qualité sont : acceptable, médiocre et inacceptable. Les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage influencent généralement la qualité des données.

Pour établir le niveau de qualité type d'une estimation, les utilisateurs devraient d'abord établir le nombre de répondants inclus dans le calcul de l'estimation. Si ce nombre est cinq ou moins, on ne devrait pas diffuser l'estimation pondérée afin de se conformer aux politiques de confidentialité. Pour les estimations pondérées reposant sur des tailles d'échantillon supérieures à cinq répondants, les utilisateurs devraient déterminer le coefficient de variation de l'estimation décrit dans le chapitre 13 et suivre les lignes directrices ci-après. On doit appliquer ces lignes directrices aux estimations arrondies pondérées. Toute estimation dont le niveau de qualité est médiocre ou inacceptable devrait être accompagnée d'une mise en garde à l'intention des utilisateurs subséquents.

Tableau 12.1

Lignes directrices concernant le niveau de qualité en fonction du CV d'une estimation particulière

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	<p>Caractéristiques des estimations : Taille de l'échantillon supérieure à cinq et coefficients de variation situés dans un intervalle de 0,0 % à 16,5 %.</p> <p>Aucune mise en garde n'est requise.</p>
2) Médiocre	<p>Caractéristiques des estimations : Taille de l'échantillon supérieure à cinq et coefficients de variation situés dans un intervalle de 16,6 % à 33,3 %.</p> <p>On doit désigner ces estimations par la lettre M (ou un autre identificateur semblable). Elles doivent faire l'objet d'une mise en garde à l'intention des utilisateurs subséquents au sujet du niveau d'erreur élevé associé aux estimations.</p>
3) Inacceptable	<p>On ne devrait pas diffuser les estimations associées à une taille de l'échantillon de cinq ou moins, afin de se conformer aux politiques de confidentialité de l'organisme.</p> <p>Pour des tailles d'échantillon supérieures à cinq mais liées à des coefficients de variation très élevés dépassant 33,3 %, Statistique Canada recommande de ne pas diffuser d'estimations, en raison de leur niveau de qualité inacceptable. Cependant, si l'utilisateur choisit de le faire, il doit alors désigner les estimations par la lettre I (ou un autre identificateur semblable) et les diffuser avec la mise en garde suivante :</p> <p>« Veuillez noter que ces estimations [marquées de la lettre I] ne répondent pas aux normes de qualité de Statistique Canada. Les conclusions tirées de ces données ne sauraient être fiables et seront fort probablement erronées. Ces données et toute conclusion qu'on pourrait en tirer ne devraient pas être publiées. Si l'utilisateur choisit de les publier, il est alors tenu de les accompagner de cette mise en garde. »</p>

13.0 Calcul de la variance

Comme nous l'avons mentionné au chapitre 10, l'EDE est une enquête par sondage et les estimations sont sujettes à l'erreur d'échantillonnage et à l'erreur non due à l'échantillonnage. L'erreur d'échantillonnage cause la variabilité; elle révèle dans quelle mesure des estimations tirées de différents échantillons possibles de même taille et de même plan, à l'aide du même estimateur, diffèrent les unes des autres. L'objet du présent chapitre est d'illustrer la mesure de la variance d'échantillonnage pour l'EDE et l'importance d'intégrer correctement le plan d'échantillonnage.

13.1 Importance de la variance

La variance d'une estimation est une bonne indication de la qualité de cette estimation. Une estimation assortie d'une variance trop élevée est considérée comme non fiable. Pour connaître le degré de variance, on utilise une mesure relative de la variabilité : le coefficient de variation (CV).

Le coefficient de variation se définit comme le ratio de la racine carrée de la variance de l'estimation (aussi désignée comme l'erreur-type) à l'estimation. Supposons que nous ayons une estimation \hat{Y} , alors le CV de cette estimation est :

$$\alpha_{\hat{Y}} = \frac{\sqrt{VAR(\hat{Y})}}{\hat{Y}}$$

L'utilisation du CV, plutôt que la variance, permet à l'analyste de comparer sur une même échelle des estimations de grandeurs différentes. On peut donc quantifier la qualité de toute estimation au moyen du CV.

13.2 Calcul de la variance et du CV d'une estimation

Afin de calculer la variance des estimations de cette enquête, l'utilisateur doit tenir compte du plan d'enquête (échantillon aléatoire stratifié) et de la correction de la non-réponse, ce qui équivaut à calculer la variance pour un plan d'échantillonnage à deux phases où la non-réponse est considérée comme la deuxième phase. On a utilisé le Système généralisé d'estimation (SGE), programme mis au point à Statistique Canada, pour calculer la variance de toutes les estimations diffusées pour cette enquête. Il convient de souligner que le texte, les tableaux et les équations renvoient à des variances et à des CV estimatifs, même si cela n'est pas mentionné explicitement.

Aucun outil n'a été mis au point pour permettre aux utilisateurs externes de calculer eux-mêmes la variance des estimations de l'EDE. Moyennant des frais, les utilisateurs désireux d'estimer les variances et les CV peuvent demander à Statistique Canada de produire des estimations de CV. Toutefois, pour donner une idée de la qualité des estimations de cette enquête, des tableaux des CV produits pour diverses estimations sont présentés ci-dessous. Les

CV présentés sont des estimations de proportions seulement. Par exemple, dans le tableau 13.1.1, le CV de la proportion de directeurs d'école ayant répondu « Peu » ou « Pas du tout » pour l'aspect « L'apprentissage des élèves » à la question 4 de la section 2 du questionnaire est de 4,8 %.

Les estimations suivantes sont fondées sur les 2 144 répondants ayant consenti à la divulgation de leurs renseignements. Les CV pour les cellules contenant un « x » ont été supprimées en raison du fait que cinq enregistrements ou moins contribuaient aux estimations ponctuelles correspondantes pour ces cellules.

Tableau 13.1.1

CV estimatifs par région – question 4 de la section 2.

« Dans quelle mesure croyez-vous que les changements observés durant la dernière décennie auront des effets positifs sur les aspects suivants au sein de votre école? »

Boîte	Aspect	Réponse	Canada	Atlantique	Qc	Ont.	Prairies	C.-B.	Territoires
									%
2245	L'apprentissage des élèves	« Peu » ou « Pas du tout »	4,8	9,5	13,2	8,7	9,7	11,0	17,9
		« Dans une large mesure » ou « Dans une certaine mesure »	1,0	2,4	1,7	2,2	2,1	2,7	5,9
2246	L'intégration des élèves à la société	« Peu » ou « Pas du tout »	3,3	7,2	11,4	5,3	6,6	8,3	13,8
		« Dans une large mesure » ou « Dans une certaine mesure »	1,5	3,3	2,0	3,6	3,2	3,5	8,1
2247	La professionnalisation des enseignants	« Peu » ou « Pas du tout »	3,1	9,6	8,4	5,6	6,8	5,7	18,5
		« Dans une large mesure » ou « Dans une certaine mesure »	1,6	2,5	2,7	3,3	3,0	5,1	7,7
2248	Vos fonctions de directeur d'école	« Peu » ou « Pas du tout »	3,8	8,9	9,5	6,6	8,1	8,4	22,5
		« Dans une large mesure » ou « Dans une certaine mesure »	1,3	2,6	2,4	2,8	2,5	3,5	6,2
2249	L'efficacité du système scolaire	« Peu » ou « Pas du tout »	3,2	7,4	7,2	6,1	6,4	7,4	14,8
		« Dans une large mesure » ou « Dans une certaine mesure »	1,6	3,1	3,1	3,1	3,1	4,0	8,9
2250	Les relations avec les parents	« Peu » ou « Pas du tout »	3,1	7,7	8,6	5,2	6,5	7,8	13,7
		« Dans une large mesure » ou « Dans une certaine mesure »	1,6	3,0	2,6	3,5	3,0	3,7	9,2
2251	La reconnaissance de la mission de l'école	« Peu » ou « Pas du tout »	3,1	7,6	10,0	5,3	5,9	7,4	14,1
		« Dans une large mesure » ou « Dans une certaine mesure »	1,5	3,1	2,2	3,5	3,4	3,9	9,3

Tableau 13.1.2

CV estimatifs par niveau d'année et par emplacement – question 4 de la section 2.

« Dans quelle mesure croyez-vous que les changements observés durant la dernière décennie auront des effets positifs sur les aspects suivants au sein de votre école? »

Boîte	Aspect	Réponse	Niveau d'année			Emplacement		%
			Canada	Primaire	Mixte	Secondaire	Rural	
2245	L'apprentissage des élèves	« Peu » ou « Pas du tout »	4,8	6,4	12,2	8,6	9,3	5,6
		« Dans une large mesure » ou « Dans une certaine mesure »	1,0	1,2	3,6	2,3	1,8	1,3
2246	L'intégration des élèves à la société	« Peu » ou « Pas du tout »	3,3	4,2	9,3	6,4	6,2	3,9
		« Dans une large mesure » ou « Dans une certaine mesure »	1,5	1,9	4,8	3,1	2,9	1,8
2247	La professionnalisation des enseignants	« Peu » ou « Pas du tout »	3,1	4,1	9,7	5,6	6,4	3,6
		« Dans une large mesure » ou « Dans une certaine mesure »	1,6	1,9	4,5	3,4	2,7	1,9
2248	Vos fonctions de directeur d'école	« Peu » ou « Pas du tout »	3,8	4,8	11,5	7,1	7,1	4,5
		« Dans une large mesure » ou « Dans une certaine mesure »	1,3	1,6	3,9	2,8	2,4	1,6
2249	L'efficacité du système scolaire	« Peu » ou « Pas du tout »	3,2	4,2	8,9	5,8	6,2	3,7
		« Dans une large mesure » ou « Dans une certaine mesure »	1,6	1,9	5,0	3,4	2,7	1,9
2250	Les relations avec les parents	« Peu » ou « Pas du tout »	3,1	4,2	8,2	5,8	5,8	3,8
		« Dans une large mesure » ou « Dans une certaine mesure »	1,6	1,8	5,3	3,5	2,9	1,9
2251	La reconnaissance de la mission de l'école	« Peu » ou « Pas du tout »	3,1	4,0	8,4	6,1	6,1	3,7
		« Dans une large mesure » ou « Dans une certaine mesure »	1,5	1,9	4,9	3,2	2,8	1,9

Tableau 13.2.1
CV estimatifs par région – question 2 de la section 3.

« Veuillez indiquer votre niveau de satisfaction à propos des points suivants : »

Boîte	Sujet	Réponse	Canada	Atlantique	Qc	Ont.	Prairies	C.-B.	Territoires
						%			
3262	L'appui de mes supérieurs	« Plutôt insatisfait » ou « Entièrement insatisfait »	5,4	11,8	11,3	12,9	10,3	11,7	29,0
		« Entièrement satisfait » ou « Plutôt satisfait »	0,9	2,0	2,1	1,7	2,0	2,6	4,8
3263	Ma charge de travail	« Plutôt insatisfait » ou « Entièrement insatisfait »	2,0	4,1	4,1	3,6	4,9	5,4	12,2
		« Entièrement satisfait » ou « Plutôt satisfait »	2,3	5,5	5,2	4,9	3,9	5,3	8,4
3264	La rémunération	« Plutôt insatisfait » ou « Entièrement insatisfait »	2,3	4,0	4,2	5,1	4,6	6,8	18,9
		« Entièrement satisfait » ou « Plutôt satisfait »	2,0	5,9	5,0	3,5	4,1	4,1	5,5
3265	La reconnaissance de ma fonction	« Plutôt insatisfait » ou « Entièrement insatisfait »	2,7	5,3	4,9	5,8	5,2	8,2	14,5
		« Entièrement satisfait » ou « Plutôt satisfait »	1,7	4,2	4,4	3,1	3,8	3,4	7,7
3266	L'incidence sur ma vie familiale	« Plutôt insatisfait » ou « Entièrement insatisfait »	1,7	3,6	3,8	3,2	3,4	4,4	10,5
		« Entièrement satisfait » ou « Plutôt satisfait »	2,9	6,6	5,9	5,6	6,0	6,6	11,6
3267	L'encadrement de mon travail par des normes juridiques	« Plutôt insatisfait » ou « Entièrement insatisfait »	2,9	5,0	7,3	5,1	5,8	8,2	14,0
		« Entièrement satisfait » ou « Plutôt satisfait »	1,8	4,8	3,2	3,7	3,5	3,8	9,9
3268	Mon imputabilité	« Plutôt insatisfait » ou « Entièrement insatisfait »	4,8	9,0	10,4	8,4	10,9	14,6	19,2
		« Entièrement satisfait » ou « Plutôt satisfait »	1,0	2,6	2,1	2,1	1,6	1,9	5,6
3269	Mon développement professionnel	« Plutôt insatisfait » ou « Entièrement insatisfait »	6,0	11,3	14,6	10,7	13,1	15,8	24,9
		« Entièrement satisfait » ou « Plutôt satisfait »	0,8	2,0	1,5	1,7	1,5	1,8	6,2
3270	Mon autonomie professionnelle	« Plutôt insatisfait » ou « Entièrement insatisfait »	4,5	9,4	11,0	7,6	11,2	11,1	14,6
		« Entièrement satisfait » ou « Plutôt satisfait »	1,0	2,4	2,0	2,3	1,7	2,6	5,7

Tableau 13.3
CV estimatifs par région – question 6 de la section 4.

« Voici différents problèmes pouvant survenir à l'école. Dans quelle mesure chacun de ces problèmes nuit-il à la bonne marche de votre école? »

Boîte	Problème	Réponse	Canada	Atlantique	Qc	Ont.	Prairies	C.-B.	Territoires
						%			
4224	Conflits entre élèves	« Peu » ou « Pas du tout »	2,0	4,5	5,4	4,1	3,6	3,5	10,2
		« Dans une large mesure » ou « Dans une certaine mesure »	2,3	5,1	4,0	4,2	5,4	7,8	9,3
4225	Intimidation entre élèves	« Peu » ou « Pas du tout »	1,9	4,5	5,2	4,1	3,4	2,8	9,1
		« Dans une large mesure » ou « Dans une certaine mesure »	2,4	5,2	4,2	4,2	5,7	9,2	10,5
4226	Problèmes de santé des élèves	« Peu » ou « Pas du tout »	1,2	2,4	2,4	2,3	2,4	2,6	7,2
		« Dans une large mesure » ou « Dans une certaine mesure »	4,2	9,7	9,2	7,8	8,4	11,0	16,3
4227	Détérioration de la situation socio-économique des familles des élèves	« Peu » ou « Pas du tout »	2,2	4,9	6,3	3,8	4,2	5,2	15,8
		« Dans une large mesure » ou « Dans une certaine mesure »	2,2	4,6	3,5	4,8	4,8	5,6	6,7
4228	Infractions contre les biens (vandalisme, vol) par les élèves	« Peu » ou « Pas du tout »	1,1	2,4	3,0	2,1	2,3	2,3	8,8
		« Dans une large mesure » ou « Dans une certaine mesure »	4,3	10,0	7,6	8,7	9,4	13,2	13,1
4229	Possession d'armes par les élèves	« Peu » ou « Pas du tout »	0,9	1,8	2,8	1,6	1,7	2,2	3,4
		« Dans une large mesure » ou « Dans une certaine mesure »	10,9	26,7	23,1	19,1	21,3	32,3	x
4230	Consommation d'alcool et de stupéfiants par les élèves	« Peu » ou « Pas du tout »	1,2	2,6	3,5	2,1	2,4	3,1	6,1
		« Dans une large mesure » ou « Dans une certaine mesure »	3,7	8,8	7,5	7,3	7,8	10,2	10,5
4231	Conduite insolente des élèves à l'endroit des enseignants	« Peu » ou « Pas du tout »	1,6	3,6	4,3	3,6	2,8	2,9	7,0
		« Dans une large mesure » ou « Dans une certaine mesure »	2,9	6,2	5,5	5,0	6,7	9,9	11,8
4232	Agressions verbales ou physiques sur un membre du personnel par les élèves	« Peu » ou « Pas du tout »	1,1	2,5	3,0	2,2	2,0	2,1	6,1
		« Dans une large mesure » ou « Dans une certaine mesure »	5,1	10,5	8,6	9,8	11,7	19,2	18,6
4233	Absentéisme des élèves	« Peu » ou « Pas du tout »	1,5	3,1	3,0	2,9	3,2	3,7	16,2
		« Dans une large mesure » ou « Dans une certaine mesure »	2,9	6,8	7,0	5,3	5,6	7,2	6,5
4234	Sexisme et harcèlement sexuel entre élèves	« Peu » ou « Pas du tout »	1,0	2,0	2,5	1,9	1,7	2,0	4,3
		« Dans une large mesure » ou « Dans une certaine mesure »	6,4	13,9	12,9	10,9	13,4	23,0	27,6

Enquête auprès des directeurs d'école, 2004-2005 – Guide de l'utilisateur

4235	Racisme entre les élèves	« Peu » ou « Pas du tout »	1,0	2,1	2,5	1,8	2,1	2,0	3,3
		« Dans une large mesure » ou « Dans une certaine mesure »	6,9	17,9	17,0	11,5	12,4	20,1	x
4236	Consommation d'alcool et de stupéfiants par des membres du personnel	« Peu » ou « Pas du tout »	0,9	1,9	2,6	1,6	1,6	2,1	3,1
		« Dans une large mesure » ou « Dans une certaine mesure »	13,5	28,3	29,5	23,6	27,7	x	x
4237	Roulement du corps enseignant	« Peu » ou « Pas du tout »	1,4	2,9	5,1	2,5	2,1	3,2	9,1
		« Dans une large mesure » ou « Dans une certaine mesure »	3,4	8,8	4,7	7,3	9,6	9,6	11,5
4238	Absentéisme des enseignants	« Peu » ou « Pas du tout »	1,1	1,9	3,6	2,1	1,9	2,6	3,6
		« Dans une large mesure » ou « Dans une certaine mesure »	4,2	13,3	6,2	9,1	10,1	11,5	21,1
4239	Chahut des élèves durant les cours	« Peu » ou « Pas du tout »	1,5	3,5	3,1	3,2	2,9	3,0	10,9
		« Dans une large mesure » ou « Dans une certaine mesure »	3,3	6,5	8,0	5,6	6,7	9,4	11,0
4240	Retard des élèves	« Peu » ou « Pas du tout »	1,4	3,1	3,1	2,9	2,7	3,5	12,9
		« Dans une large mesure » ou « Dans une certaine mesure »	3,2	7,5	7,8	5,7	6,7	8,2	7,2
4241	Décrochage scolaire	« Peu » ou « Pas du tout »	1,4	2,4	3,7	2,9	2,5	2,8	6,1
		« Dans une large mesure » ou « Dans une certaine mesure »	5,2	16,7	8,3	11,5	10,1	19,2	10,3
4242	Apathie des élèves	« Peu » ou « Pas du tout »	1,5	3,0	4,1	2,8	2,9	3,1	8,0
		« Dans une large mesure » ou « Dans une certaine mesure »	3,2	7,2	5,7	6,5	6,6	9,5	8,9
4243	Conflits entre parents et enseignants	« Peu » ou « Pas du tout »	1,1	2,1	2,8	2,1	2,3	2,3	4,4
		« Dans une large mesure » ou « Dans une certaine mesure »	4,4	11,5	8,6	8,2	8,4	12,3	20,6
4244	Plaintes des parents et des élèves	« Peu » ou « Pas du tout »	1,1	2,1	3,2	2,3	2,1	2,1	3,6
		« Dans une large mesure » ou « Dans une certaine mesure »	4,2	11,3	7,3	7,8	8,8	12,6	19,0

En général, on doit appliquer les règles suivantes lorsqu'on calcule et utilise les CV.

Règle 1 : Estimations des proportions ou des pourcentages de directeurs d'école présentant une caractéristique

Les CV calculés dans les tableaux qui précèdent le sont pour des proportions. Ils peuvent donc être utilisés tels qu'ils apparaissent dans le chiffrier. Pour d'autres estimations semblables de proportions (par ex., pour d'autres questions), les CV des tableaux ci-dessus pourraient servir à titre indicatif, mais l'utilisateur devrait communiquer avec Statistique Canada pour obtenir les CV calculés pour ces caractéristiques spécifiques. Veuillez noter que toutes les estimations de CV pour des proportions ne sont valides que si la non-réponse à une question est faible.

Règle 2 : Estimations du nombre de directeurs d'école présentant une caractéristique (agrégats)

Le CV estimatif du nombre de directeurs d'école présentant une caractéristique est, en théorie, identique à celui de la proportion qu'il représente (la proportion correspondante de directeurs d'école présentant une caractéristique) et très proche, mais légèrement supérieur, si la non-réponse à une question est faible. On ne doit pas utiliser ce rapport lorsque la non-réponse à la question concernant la caractéristique est élevée et (ou) que la taille du domaine d'intérêt est relativement faible.

Par exemple, supposons que nous ayons une estimation $\hat{Y} = 1\ 000$ directeurs présentant une certaine caractéristique sur le total estimatif $\hat{X} = 5\ 000$ directeurs du domaine d'intérêt; le CV de \hat{Y} devrait alors être proche du CV de la proportion $\frac{\hat{Y}}{\hat{X}}$.

Règle 3 : Estimations des différences entre des agrégats, des pourcentages et des ratios

L'erreur-type d'une différence entre deux estimations est à peu près égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. Autrement dit,

l'approximation de l'erreur-type de différence $\hat{d} = \hat{X}_1 - \hat{X}_2$ est :

$$\hat{\sigma}_{\hat{d}} = \sqrt{(\hat{X}_1 \hat{\alpha}_1)^2 + (\hat{X}_2 \hat{\alpha}_2)^2}$$

où $\hat{\alpha}_1$ et $\hat{\alpha}_2$ sont respectivement les CV de \hat{X}_1 et \hat{X}_2 . On obtient le CV de \hat{d} en divisant son erreur-type par \hat{d} elle-même :

$$\hat{\alpha}_{\hat{d}} = \frac{\hat{\sigma}_{\hat{d}}}{\hat{d}}$$

Il s'agit d'une bonne approximation du CV d'une différence entre des caractéristiques distinctes et non corrélées, mais elle a tendance à surestimer le CV si \hat{X}_1 et \hat{X}_2 sont corrélées positivement et à le sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélées négativement. Moyennant des frais, on peut obtenir les estimations de CV pour les valeurs \hat{X}_1 et \hat{X}_2 corrélées en s'adressant à Statistique Canada.

Règle 4 : Estimations de ratios

L'erreur-type du ratio des estimations est approximativement égale à la racine carrée de la somme des carrés de chaque CV pris séparément, multipliée par le ratio lui-même. Le CV d'un tel ratio $\hat{R} = \hat{X}_1 / \hat{X}_2$ est :

$$\hat{\alpha}_{\hat{R}} = \sqrt{\hat{\alpha}_1^2 + \hat{\alpha}_2^2}$$

où $\hat{\alpha}_1$ et $\hat{\alpha}_2$ sont les CV respectifs de \hat{X}_1 et de \hat{X}_2 respectivement. Comme dans l'estimation du CV de la règle 3, cette estimation tend à surestimer le CV si \hat{X}_1 et \hat{X}_2 sont corrélées positivement et à le sous-estimer si \hat{X}_1 et \hat{X}_2 sont corrélées négativement.

Règle 5 : Estimations des différences entre des ratios

Il s'agit ici simplement de combiner les règles 3 et 4. On commence par calculer le CV des deux ratios au moyen de la règle 4, puis on trouve le CV de leur différence par la règle 3.

13.3 Calcul des limites de confiance à l'aide du coefficient de variation

Bien que les coefficients de variation soient largement utilisés, l'intervalle de confiance d'une estimation donne une mesure intuitive plus significative de l'erreur d'échantillonnage. Un intervalle de confiance est le niveau de confiance que la valeur vraie de la population se situe dans une fourchette donnée de valeurs. L'intervalle de confiance à 95 % signifie que si l'enquête se répète à l'infini avec le même échantillonnage et le même taux de non-réponse, 95 % des ensembles de réponses produiront alors des estimations dont l'intervalle de confiance contiendrait la vraie valeur.

Les intervalles de confiance pour une estimation, \hat{X} , sont généralement exprimés sous forme de deux nombres, l'un inférieur et l'autre supérieur à l'estimation, comme dans $(\hat{X} \pm k)$, où k varie selon le niveau de confiance souhaité et l'erreur-type de l'estimation. Comme la véritable erreur-type des estimations est inconnue, les intervalles de confiance sont estimés en fonction de l'ensemble des répondants.

Une fois déterminée l'erreur-type d'une estimation, on peut calculer des intervalles de confiance pour les estimations en partant de l'hypothèse qu'avec un échantillon et une population suffisamment vastes et une différence suffisante entre ces deux nombres (N-n), les diverses estimations obtenues pour une caractéristique de la population sont réparties selon une distribution normale autour de la vraie valeur de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que la différence entre une estimation de l'échantillon et la vraie valeur pour la population soit inférieure à une erreur-type, environ 95 chances sur 100 que la différence soit inférieure à deux erreurs-types et environ 99 chances sur 100 que la différence soit inférieure à trois erreurs-types. On appelle ces différents degrés de confiance des niveaux de confiance.

Si l'utilisateur veut déterminer d'autres intervalles de confiance, il peut employer la formule ci-après pour obtenir l'intervalle de confiance (IC \hat{X}) :

$$IC \hat{X} = (\hat{X} \pm z\hat{\alpha}_{\hat{X}})$$

où $\hat{\alpha}_{\hat{X}}$ est le CV trouvé pour \hat{X} et

$z = 1$ si l'on désire un intervalle de confiance à 68 %;

$z = 1,65$ si l'on désire un intervalle de confiance à 90 %;

$z = 1,96$ si l'on désire un intervalle de confiance à 95 %;

$z = 2,58$ si l'on désire un intervalle de confiance à 99 %.

Exemple d'utilisation du coefficient de variation pour obtenir des limites de confiance

L'exemple suivant renvoie à la question 4 de la section 2 (voir le tableau 13.1.1). Voici comment on calculerait un intervalle de confiance à 95 % pour la proportion estimée de directeurs d'école canadiens qui croient que les changements observés durant la dernière décennie auront peu ou pas du tout d'effets positifs sur l'apprentissage des élèves :

$$\hat{X} = 17,4 \% \text{ (ou, exprimé comme une proportion, } 0,174)$$

$$z = 1,96$$

Le coefficient de variation de \hat{X} est 4,8 % (ou 0,048 s'il est exprimé comme une proportion).

$$IC \hat{X} = \{0,174 - (1,96) (0,174) (0,048), 0,174 + (1,96) (0,174) (0,048)\}$$

$$IC \hat{X} = \{0,174 - 0,016, 0,174 + 0,016\}$$

$$IC \hat{X} = \{0,158, 0,190\}$$

On peut donc dire, avec un niveau de confiance à 95 %, qu'entre 15,8 % et 19,0 % des directeurs d'école à l'échelle nationale croient que les changements observés durant la dernière décennie auront peu ou pas du tout d'effets positifs sur l'apprentissage des élèves.

13.4 Tests d'hypothèses

On peut aussi utiliser les erreurs-types pour vérifier des hypothèses; cette technique permet de faire des inférences au sujet de différences possibles entre les paramètres d'une population à l'aide d'estimations basées sur un échantillon. Ces estimations peuvent être des nombres, des moyennes, des pourcentages, des ratios, etc. Les tests peuvent être effectués à divers niveaux de signification, un niveau de signification étant la probabilité de conclure que les caractéristiques sont différentes alors qu'en fait, elles sont identiques.

Posons \hat{X}_1 et \hat{X}_2 comme des estimations d'échantillon pour deux caractéristiques d'intérêt.

Comme dans la règle 3 à la section 13.2, supposons que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$

est $\hat{\sigma}_d$. Si la quantité $(\hat{X}_1 - \hat{X}_2) / \hat{\sigma}_d$ se situe entre -1,96 et +1,96, il n'y a pas de différence

significative entre les caractéristiques au niveau de signification de 5 %. Toutefois, si ce

rapport est inférieur à -1,96 ou supérieur à +1,96, la différence observée est significative au

niveau de 5 %. Comme l'indique la règle 3 à la section 13.2, cette méthode est fiable

uniquement si elle s'applique à deux caractéristiques d'intérêt non corrélées; autrement, la

quantité $(\hat{X}_1 - \hat{X}_2) / \hat{\sigma}_d$ peut s'avérer inexacte.

On peut aussi déterminer si deux estimations sont significativement différentes en comparant leurs intervalles de confiance. Lorsque deux intervalles de confiance se chevauchent, on peut conclure qu'il n'y a pas de différence significative entre les estimations correspondantes (même si, en théorie, il pourrait y avoir une différence significative dans ce cas). Il convient de souligner que cette méthode est plutôt approximative, car elle revient à calculer approximativement l'intervalle de confiance pour la différence des deux estimations. Par conséquent, l'approximation de l'intervalle de confiance sera légèrement supérieure à celle qui est dérivée dans la section 13.3. Pour cette raison, le recours à cette règle empirique prudente entraîne un faible risque de ne pas détecter certains écarts significatifs (mais très minces) selon le niveau de confiance utilisé.

13.5 Coefficients de variation d'estimations quantitatives

Seules quelques variables quantitatives ont pu être estimées à partir de cette enquête (voir la définition d'une estimation quantitative à la section 12.3.2). Seules les questions 8 à 15 de la section 1 ont pu être utilisées pour calculer des estimations quantitatives. Toutefois, aucune analyse ne devrait être effectuée sur la base des questions 9 et 14 (pour plus de détails, voir la section 10.6).

Le coefficient de variation d'un total quantitatif sera généralement supérieur à celui de l'estimation nominale correspondante. Si l'estimation nominale correspondante ne peut être diffusée, il en sera de même pour l'estimation quantitative. Par exemple, si l'on tient compte uniquement des directeurs d'école primaire, le CV du nombre moyen d'années d'expérience à titre d'enseignant au primaire serait sans doute supérieur au CV de la proportion de directeurs d'école possédant une expérience d'enseignement à l'école primaire. Par conséquent, si le coefficient de variation de la proportion ne peut être diffusé, ce sera aussi le cas pour le coefficient de variation de l'estimation quantitative correspondante.

Annexe A : Questionnaire

Cette annexe présente le questionnaire de l'EDE envoyé aux directeurs d'école.

Le questionnaire reproduit en annexe est en français. Veuillez vous reporter à la version anglaise du Guide de l'utilisateur pour obtenir un exemplaire du questionnaire en anglais.

Annexe B : Caractéristiques d'une école

Cette annexe présente une vue d'ensemble de définitions utilisées pour dériver les variables des caractéristiques d'une école.

Niveau d'enseignement de l'école – Écoles primaires, secondaires et primaires-secondaires mixtes

Les écoles au Canada sont classées comme écoles **primaires** si elles offrent un enseignement de la première à la sixième année ou une majorité des années du niveau primaire; comme écoles **secondaires** si elles offrent un enseignement à partir de la septième année ou une majorité des années du niveau secondaire, et comme **écoles mixtes primaires-secondaires** si elles offrent une combinaison d'années des niveaux primaire et secondaire.

Emplacement de l'école : Écoles urbaines et rurales

Les écoles **rurales** s'entendent des établissements situés en milieu rural, dans les petites villes de même que dans les banlieues rurales d'une agglomération plus importante (régions métropolitaines de recensement (RMR) ou agglomérations de recensement (AR)). Les écoles **urbaines** s'entendent des établissements situés dans une RMR ou une AR, mais non dans les banlieues rurales.

Financement de l'école – Financement public et privé

On a classé les établissements en écoles publiques ou privées selon que la prise de décision quant aux affaires de l'établissement relève, en dernière instance, d'un organisme public ou privé. Sont classés dans la catégorie des écoles **publiques** les établissements qui, selon les déclarations du directeur, sont : contrôlés ou administrés directement par une autorité scolaire ou un organisme public; ou contrôlés et administrés soit directement par un organisme public, soit par un organe directeur (conseil, comité, etc.), dont la majorité des membres est nommée par des pouvoirs publics ou élus par le public. Sont classés dans la catégorie des écoles **privées** les établissements qui, selon les déclarations du directeur, sont contrôlés et administrés par un organisme non gouvernemental (p. ex., une église, un syndicat ou une entreprise) ou dont le conseil d'administration est majoritairement formé de membres non nommés par un organisme public.

Taille de l'école – Petite, moyenne et grande

La taille de l'école est déterminée en fonction de la répartition du nombre d'élèves inscrits dans des écoles primaires, secondaires et mixtes. Les écoles dont le nombre d'élèves se situe dans le tiers inférieur de la distribution ($<33,3\%$) sont considérées comme de « **petites écoles** ». Les « **écoles moyennes** » sont celles dont le nombre d'élèves se situe dans le tiers intermédiaire de la distribution ($\leq 33,3\%$ et $\leq 66,6\%$). Les « **grandes écoles** » s'entendent des établissements dont le nombre d'élèves se situe dans le tiers supérieur de la distribution ($>66,6\%$). La taille des écoles est donc définie comme suit :

	Primaire	Secondaire	Mixte primaire et secondaire
Petite	Moins de 200 élèves	Moins de 300 élèves	Moins de 60 élèves
Moyenne	De 200 à 350 élèves	De 300 à 700 élèves	De 60 à 200 élèves
Grande	Plus de 350 élèves	Plus de 700 élèves	Plus de 200 élèves