



Guide de l'utilisateur des microdonnées

Enquête auprès des titulaires d'un doctorat 2003-2004

Table des matières

1.	Introduction.....	5
2.	Contexte.....	6
3.	Objectifs.....	7
4.	Concepts et définitions.....	8
4.1	Contenu.....	8
4.2	Concepts et définitions.....	9
5.	Méthodologie de l'enquête.....	10
5.1	Population visée.....	10
5.2	Plan de sondage.....	10
6.	Collecte des données.....	11
6.1	Distribution du matériel d'enquête, collecte des données et listes.....	11
6.2	Supervision et contrôle de qualité.....	13
6.3	Non-réponses.....	13
7.	Traitement des données.....	14
7.1	Saisie des données.....	14
7.2	Vérification.....	14
7.3	Codage des questions ouvertes.....	15
7.3.1	Codage du champ du domaine d'études de l'ETD.....	15
7.3.2	Codage de l'industrie et de la branche d'activité.....	15
7.3.3	Codage des réponses « Autre – Veuillez préciser ».....	15
7.4	Création de variables dérivées.....	16
7.5	Pondération.....	17
8.	Qualité des données.....	18
8.1	Taux de réponse.....	18
8.2	Erreurs relatives à l'enquête.....	18
8.2.1	Base de sondage.....	18
8.2.2	Collecte des données.....	19
8.2.3	Traitement des données.....	19
8.2.4	Non-réponse.....	19
9.	Lignes directrices pour la totalisation, l'analyse et la diffusion des données.....	22
9.1	Lignes directrices sur l'arrondissement.....	22
9.2	Lignes directrices pour la pondération d'une totalisation.....	23
9.3	Définitions des types d'estimations : catégoriques et quantitatives.....	23
9.3.1	Estimations catégoriques.....	23
9.3.2	Estimations quantitatives.....	23
9.3.3	Totalisation d'estimations catégoriques.....	24
9.3.4	Totalisation d'estimations quantitatives.....	24
9.4	Lignes directrices pour l'analyse statistique.....	25
9.5	Lignes directrices sur la diffusion de coefficients de variation.....	26
9.6	Seuils pour la diffusion des estimations pour l'Enquête de 2003-2004 auprès des titulaires d'un doctorat.....	28

10.	Tableau de variabilité approximative	29
10.1	Comment utiliser le tableau de coefficients de variation pour des estimations catégoriques.....	29
10.1.1	Exemples d'utilisation du tableau de coefficients de variation pour des estimations catégoriques	31
10.2	Comment utiliser le tableau des coefficients de variation pour obtenir des limites de confiance.....	35
10.2.1	Exemple d'utilisation du tableau de coefficients de variation pour obtenir les limites de confiance.....	36
10.3	Comment utiliser le tableau de coefficients de variation pour effectuer un test t	36
10.3.1	Exemple d'utilisation du tableau de coefficients de variation pour effectuer un test t	37
10.4	Coefficients de variation pour des estimations quantitatives	37
10.5	Tableau de variabilité approximative	38
11.	Questionnaire, dictionnaire de données et feuilles de codage	39
12.	Cliché d'enregistrement à valeurs univariées	40

1. Introduction

L'Enquête de 2003-2004 auprès des titulaires d'un doctorat (ETD) a été réalisée par Statistique Canada entre septembre 2003 et septembre 2004 avec la collaboration de tous les établissements canadiens d'enseignement qui décernent un diplôme de doctorat. Ce manuel a été conçu pour faciliter la manipulation des fichiers de microdonnées des résultats de l'enquête.

Les questions au sujet de l'ensemble de données ou de son utilisation doivent être adressées à :

Statistique Canada
Services aux clients
Centre de la statistique de l'éducation
Pièce SC-2200, immeuble Principal
Pré Tunney
Ottawa (Ontario) K1A 0T6

Téléphone : (613) 951-7608 ou sans frais 1 800 307-3382

Télécopieur : (613) 951-9040

Adresse électronique : educationstats@statcan.ca

2. Contexte

L'Enquête auprès des titulaires d'un doctorat (ETD) a été conçue afin de recueillir de l'information intégrée sur les caractéristiques et les projets des titulaires d'un doctorat. L'ETD est un véhicule pour ceux et celles qui voudront faire de nouvelles recherches et analyser la relation entre le parcours éducatif des diplômés et leurs futurs projets de carrière ou d'études.

L'enquête sert à recueillir des données sur le cheminement postsecondaire du diplômé, les caractéristiques de ses programmes d'études, ses sources de financement pendant ses études supérieures, son niveau d'endettement, ainsi que ses projets de carrière et de mobilité après l'obtention de son diplôme.

3. Objectifs

L'Enquête auprès des titulaires d'un doctorat (ETD) est un recensement annuel des diplômés du doctorat mené pour la première fois à l'échelle du Canada au cours de l'année universitaire 2003-2004. Son objectif principal consiste à recueillir des données sur les titulaires d'un doctorat au Canada afin d'informer le gouvernement, les associations, les universités et d'autres intervenants des caractéristiques et des projets de ces diplômés hautement qualifiés une fois terminé leur programme de doctorat.

Ces données sont importantes pour améliorer l'enseignement supérieur puisqu'elles fourniront aux agences gouvernementales et privées l'information nécessaire pour élaborer des programmes et prendre des décisions stratégiques. Des données au sujet des diplômés de chaque établissement sont également fournies aux bureaux de recherche des établissements ayant participé à l'enquête pour qu'ils les utilisent.

Principaux objectifs de l'enquête :

- Mesurer l'impact des diverses sources de financement institutionnel.
- Recueillir de l'information concernant la rétention des étudiants de doctorat au Canada.
- Mieux comprendre le système de financement des études supérieures et le niveau d'endettement des étudiants.
- Permettre aux planificateurs du marché du travail d'évaluer l'apport de ressources humaines hautement qualifiées dans les divers domaines du bassin de main-d'oeuvre.
- Permettre une analyse du cheminement jusqu'à l'obtention du diplôme de doctorat, et de l'impact des étudiants étrangers.

4. Concepts et définitions

4.1 Contenu

Cette section présente les divers champs d'interrogation de l'Enquête auprès des titulaires d'un doctorat.

Section	Contenu
PARTIE A – ÉDUCTION	<p>Titre de la thèse, domaine d'études et numéro du domaine.</p> <p>Sources de financement pendant les études postsecondaires et montant de la dette.</p> <p>Information détaillée sur les diplômes universitaires dont le répondant est titulaire ainsi que sur tout autre diplôme ou certificat postsecondaire.</p>
PARTIE B – PROJETS APRÈS L'OBTENTION DU DIPLÔME	<p>Information sur les projets du diplômé et le pays où il a l'intention de résider; dans quelle mesure ces projets sont arrêtés.</p> <p>Activité principale du répondant après l'obtention de son diplôme.</p> <p>Information sur l'activité sur le marché du travail du diplômé, par exemple profession, secteur de l'industrie.</p> <p>Salaire espéré ou actuel après l'obtention du diplôme ou source de financement des études ou recherches postdoctorales.</p>
PARTIE C – RENSEIGNEMENTS DE BASE	<p>Information de base comme l'état matrimonial, le nombre d'enfants à charge, le niveau de scolarité des parents, la nationalité et l'origine ethnoculturelle.</p>
CONTACTS	<p>Numéro de téléphone et adresse électronique du répondant, plus le nom et l'adresse d'une relation pour fins de suivi.</p>
ENTENTE DE COMMUNICATION DES DONNÉES	<p>Demande au diplômé l'autorisation de communiquer les réponses à l'établissement où il a fait son doctorat.</p>

4.2 Concepts et définitions

Doctorat axé sur la recherche vs doctorat non axé sur la recherche

L'ETD porte sur tous les titulaires d'un doctorat, qu'il soit axé sur la recherche ou non. De façon générale, un doctorat est axé sur la recherche quand l'étudiant doit rédiger un mémoire ou une thèse, ou un projet de travail original équivalent (p. ex., une composition musicale) et si l'objectif exclusif du diplôme n'est pas de permettre l'accès à la pratique d'une profession. Au Canada, tous les diplômes de doctorat sont inclus; autrement dit, les diplômes professionnels comme ceux de M.D., D.M.D., et les autres doctorats comme celui en théologie (S.T.D.), en musique (D.Mus.), etc., sont également inclus. La question A1b a été ajoutée afin de pouvoir faire des comparaisons avec l'Enquête auprès des titulaires d'un doctorat menée aux États-Unis, mais qui ne recueille de l'information que sur les doctorats de *recherche*.

Temps nécessaire pour obtenir un doctorat

Le temps nécessaire pour obtenir un doctorat s'entend du temps écoulé entre la date où le répondant commence ses études doctorales et celle où il obtient son diplôme, comme l'indique la réponse à la question A8, que le diplômé ou la diplômée ait ou non interrompu ses études pendant cette période. La variable DV_TIMPH du manuel de codes représente le temps total nécessaire pour terminer les études de doctorat.

Sources de soutien financier

Dans le questionnaire, deux questions permettent de recueillir de l'information sur les sources de soutien financier du répondant ou de la répondante pendant ses études supérieures. À la question (A5), on lui demande d'indiquer TOUTES ses sources de soutien financier parmi une liste de 28 réponses. À la question (A6), il est prié d'indiquer, parmi les sources mentionnées à la question A5, quelle était la principale source de soutien (la plus importante) et la source secondaire de soutien (deuxième en importance).

Projets arrêtés ou non arrêtés après l'obtention du diplôme

Le questionnaire comporte un certain nombre de questions sur les projets du répondant après l'obtention du diplôme (section B). C'est la réponse à la question B3 qui permet de savoir si ces projets sont arrêtés ou non. Ils le sont si la réponse est (1) « Je conserverai ou retournerai à l'emploi que j'occupais auparavant. » ou (2) « J'ai signé un contrat ou je me suis engagé à entreprendre un autre travail ou d'autres études. » La variable DV_DEFPL (les projets après l'obtention du diplôme sont arrêtés) est dérivée de la question B3.

5. Méthodologie de l'enquête

L'Enquête de 2004 auprès des titulaires d'un doctorat (ETD) a été conçue pour recueillir de l'information sur toutes les personnes qui ont obtenu un doctorat entre le 1^{er} juillet 2003 et le 30 juin 2004.

5.1 Population visée

La population de l'enquête est constituée de tous les titulaires d'un doctorat des établissements canadiens d'enseignement supérieur qui ont obtenu leur diplôme pendant la période de référence du 1^{er} juillet 2003 au 30 juin 2004. Sont spécifiquement exclus les diplômés des établissements n'ayant pas participé à l'enquête pendant la période de référence.

5.2 Plan de sondage

L'ETD a été conçue comme un recensement; partant, les questionnaires ont été distribués à tous les répondants éligibles des établissements participants.

6. Collecte des données

La collecte des données s'est effectuée de septembre 2003 à septembre 2004. Certaines parties du processus de collecte ont été prises en charge par les établissements participants, les autres étant assurées par Statistique Canada.

6.1 Distribution du matériel d'enquête, collecte des données et listes

L'enquête comportait trois volets, la distribution du matériel (une lettre de présentation, un questionnaire en format papier et une enveloppe de retour préaffranchie), la collecte des données, et la liste des diplômés.

Cohorte de l'été

En septembre 2003, chaque établissement a envoyé une trousse contenant le matériel d'enquête à ses étudiants ayant obtenu leur doctorat entre juillet 2003 et septembre 2003. Certains établissements leur ont demandé de renvoyer le questionnaire rempli à Statistique Canada au moyen de l'enveloppe de retour, tandis que d'autres les ont priés de le leur remettre dûment scellé dans l'enveloppe de retour; les autres établissements offraient les deux options.

Les établissements devaient également transmettre à Statistique Canada une liste des diplômés recevant le questionnaire, liste qui lui servirait à faire un suivi téléphonique auprès des retardataires. Voici l'information présentée dans ces listes, lorsqu'elle était disponible :

- Nom
- Adresse actuelle
- Téléphone, domicile et travail
- Sexe
- Langue de choix pour les communications
- Date d'envoi du questionnaire

Les listes étaient transmises à Statistique Canada par diverses méthodes sécuritaires.

Cohorte d'octobre 2003 à juin 2004

Les procédures de collecte des données de cette étape sont similaires à celles de l'été, à l'exception du fait que l'on demandait aux établissements :

- de distribuer le matériel d'enquête aux étudiantes et aux étudiants lorsqu'ils ou elles présentent leur demande de doctorat ou quand le diplôme leur est décerné.
- d'envoyer chaque mois à Statistique Canada une liste des diplômés ayant reçu le matériel d'enquête.

Fin du cycle de l'enquête

À la fin du cycle de l'enquête, chaque établissement devait fournir à Statistique Canada une liste complète finale de tous les étudiants ayant obtenu leur diplôme pendant la période de référence. Les listes finales de tous les établissements participants permettaient d'établir la base de sondage. Tout dépendant des procédures adoptées par l'établissement, il pouvait arriver que la liste finale ne comporte pas le nom de tous les étudiants figurant sur les listes (mensuelles) précédentes. Par exemple, certains établissements remettaient le questionnaire d'enquête de

l'ETD aux doctorants au moment de leur soutenance de thèse. C'est à ce moment-là que l'établissement transmettait la liste mensuelle de ceux et celles qui avaient reçu le questionnaire. Cependant, certains étudiants ne présentaient pas leur thèse définitive et n'obtenaient pas leur diplôme avant le cycle d'enquête suivant. On les considérait alors comme étant hors du champ de l'enquête de 2003-2004. De plus, les questionnaires reçus pendant la période de collecte, mais qui n'appartenaient pas à un diplômé figurant sur la liste finale étaient eux aussi considérés comme étant hors du champ de l'enquête.

La liste finale des diplômés ne comportait pas toute l'information-contact contenue dans les listes mensuelles, mais devait inclure spécifiquement d'autres renseignements (les champs marqués d'un astérisque sont obligatoires):

- *nom au complet du diplômé
- *nom du programme
- domaine d'études principal
- *date d'obtention du diplôme
- *sexe
- *date de naissance
- statut d'immigrant
- citoyenneté
- pays de naissance
- étudiant titulaire d'un visa ou étranger
- étudiant autochtone ou appartenant à une minorité visible
- langue maternelle

Lors de l'enquête de 2003-2004, la majorité des établissements n'ont pas transmis les renseignements non obligatoires.

6.2 Supervision et contrôle de qualité

Les questionnaires ont été remis aux diplômés par les établissements participants, qui ont reçu un manuel décrivant les concepts et les procédures de collecte de l'ETD. Une fois rempli et scellé par l'étudiant dans l'enveloppe de retour de SC, le questionnaire ne pouvait être ouvert que par Statistique Canada, qu'il ait été remis à l'établissement ou envoyé directement par le répondant. Le personnel de collecte de Statistique Canada, ayant une bonne connaissance des concepts et des procédures de l'ETD, faisait des entrevues de suivi. Ce personnel était sous la surveillance d'un agent opérationnel principal, qui vérifiait périodiquement les interviews de suivi et examinait les documents complets. Cet agent, à son tour, travaillait avec le personnel affecté au sujet au Centre de la statistique de l'éducation afin de régler les problèmes en suspens et de surveiller le processus de collecte des données.

6.3 Non-réponses

Le personnel opérationnel avait pour directive de faire tous les efforts raisonnables pour que les diplômés remettent des questionnaires complets. On cessait le suivi après cinq tentatives infructueuses, à moins que des dispositions particulières n'aient été déjà prises auparavant, dispositions que l'on respectait alors. Dans les cas où l'appel survenait à un moment inopportun, on fixait un rendez-vous avec le répondant afin de le rappeler à un moment qui lui convenait.

Cas de rejet au contrôle

Certains éléments du questionnaire revêtaient une importance cruciale pour la recherche et l'analyse. Dans le cas où le répondant avait laissé en blanc l'un de ces champs, le personnel opérationnel le contactait afin d'essayer d'obtenir l'information manquante.

En cas de non-réponse pour cause d'information sur les contacts incorrecte ou obsolète, on essayait de retracer le répondant, tant au Canada qu'à l'étranger. On s'efforçait de communiquer avec les répondants retrouvés s'ils vivaient dans un fuseau horaire où il était possible de faire un appel téléphonique pendant les heures de bureau du personnel opérationnel.

7. Traitement des données

Le principal produit de l'Enquête auprès des titulaires d'un doctorat est un fichier de microdonnées des réponses au questionnaire, ainsi que des variables dérivées et pondérées. On produit également un fichier sur l'ensemble de la population afin de permettre au logiciel de l'utilisateur de faire une estimation de l'erreur liée aux non-réponses.

7.1 Saisie des données

La saisie des données était effectuée au moyen d'ordinateurs situés au bureau central de Statistique Canada. Pendant le processus, les réponses de tous les questionnaires complets étaient codées dans une base de données. Des contrôles de valeur/étendue étaient inclus dans le programme afin de prévenir les erreurs et d'assurer l'intégrité des données.

Codage et prétraitement manuels - Cette première étape consistait à préparer les questionnaires complets en vue de la saisie des données. Chacun faisait l'objet d'un examen minutieux quant à l'exactitude des réponses et afin de s'assurer qu'il était complet. Au besoin, on attribuait des codes, que l'on notait sur le questionnaire. Les réponses devaient être lisibles afin de permettre leur enregistrement, et on les rendait plus claires si nécessaire.

Enregistrement direct des données - La saisie des données était faite par le moyen traditionnel de l'enregistrement par ordinateur. On procédait également à une vérification des données enregistrées en demandant à un autre opérateur de saisir une deuxième fois les données non textuelles de tous les questionnaires. Le contrôle de la qualité était assuré par une comparaison des deux jeux de données enregistrées et du questionnaire papier.

7.2 Vérification

La phase de vérification consiste à relever les valeurs erronées ou incohérentes dans les données de l'enquête et à apporter les modifications nécessaires.

Le premier type d'erreur est l'incohérence dans la logique du questionnaire. Les questionnaires comportant des incohérences dans l'information sont parfois l'indice que le répondant a répondu à une section qui, en fait, ne s'applique pas à lui. Dans ce cas, on élimine les données superflues. On retrouve ce type d'erreur lorsque, par exemple, un répondant indique à la question B3 qu'il n'a pas de projet de travail ou d'études d'ici un an, mais qu'il répond ensuite à la question B4 alors qu'il n'aurait pas dû.

Le deuxième type d'erreur que l'on corrige se produit lors de la vérification des fichiers d'enquête en fonction de règles de vérification préétablies pour relever les incohérences dans la logique des données. Dans ce cas, on a conçu une stratégie afin d'éliminer le moins possible de données, en établissant des priorités de façon à éliminer les valeurs les plus susceptibles d'être erronées. C'est ce qui se produit, par exemple, si un répondant se déclare canadien de naissance à la question C8, mais également immigrant reçu à la question C9 et qu'il donne la date où il est devenu immigrant reçu à la question C10. Dans ce cas, on estime que les réponses aux questions C9 et C10 sont vraisemblablement correctes et on recode la réponse à la question C8.

Le troisième type d'erreur est lié au codage « Non déclaré » appliqué lorsque le répondant ne fournit aucune réponse à une question à laquelle il aurait dû répondre.

7.3 Codage des questions ouvertes

7.3.1 Codage du champ du domaine d'études de l'ETD

Tous les domaines d'études indiqués dans le questionnaire (questions A2 et A8) sont codés en fonction de la Classification des programmes d'enseignement (CPE - 2000), en plus des codes des domaines d'études de l'ETD fournis à la page 8 du questionnaire. La plupart des domaines reçoivent un codage à six chiffres. Voir l'annexe A pour plus de renseignements sur les jeux de codes.

7.3.2 Codage de l'industrie et de la branche d'activité

Dans le cas des répondants indiquant qu'ils auraient un emploi dès l'obtention de leur diplôme, le questionnaire recueille de l'information sur le genre d'entreprise, d'industrie ou de service dans lequel opère l'employeur, le type de travail et les activités principales du répondant dans son emploi (questions B5 à B7). Avec cette information, on peut attribuer un code à l'industrie ou à la branche d'activité au moyen du Système de classification des industries de l'Amérique du Nord (SCIAN) 2002 et de la Classification nationale des professions pour statistiques (CNP-S) 2001. Voir les annexes B et C pour plus de renseignements sur le jeu de codes.

7.3.3 Codage des réponses « Autre – Veuillez préciser »

On fait un codage des réponses aux questions « Autre - Veuillez préciser » dans le cas des questions qui présentaient une liste de catégories de réponses dont la dernière était « Autre - Veuillez préciser ». Ce type de réponse en toutes lettres est possible lorsque le répondant veut donner une information qui ne fait pas déjà partie de la liste présentée, et qu'on lui donne la possibilité de donner une réponse différente. Si la réponse en toutes lettres est liée à l'une des catégories présentées, on la recode dans la catégorie appropriée. Les réponses qui ne peuvent être codées dans l'une des catégories existantes sont considérées comme appartenant réellement à la catégorie « Autre - Veuillez préciser ».

7.4 Création de variables dérivées

Un certain nombre des données des fichiers de microdonnées sont dérivées de l'information recueillie grâce au questionnaire. Dans certains cas ces variables dérivées sont évidentes et il suffit de regrouper des catégories. Dans d'autres, deux variables ou plus sont combinées pour en créer une nouvelle. Voici une liste des variables dérivées de l'ETD.

Variable	Description
DV_AGEPH	Âge au moment de l'obtention du diplôme (doctorat)
DV_MARST	État matrimonial recodé
DV_CITIZ	Citoyenneté recodée
DV_TIMPH	Temps nécessaire pour terminer les études doctorales (doctorat seulement)
DV_ACTLM	Limitation d'activité
DV_DEFPL	Projets arrêtés après l'obtention du diplôme
DV_EDUMO	Niveau de scolarité recodé de la mère
DV_EDUFA	Niveau de scolarité recodé du père
DV_DIP1	Premier degré, diplôme ou certificat postsecondaire supplémentaire obtenu
DV_DIP2	Deuxième degré, diplôme ou certificat postsecondaire supplémentaire obtenu
DV_DIP3	Troisième degré, diplôme ou certificat postsecondaire supplémentaire obtenu
DV_DIP4	Quatrième degré, diplôme ou certificat postsecondaire supplémentaire obtenu

7.5 Pondération

Bien que l'Enquête auprès des titulaires d'un doctorat soit conçue comme un recensement, on a calculé des facteurs de pondération pour tenir compte des non-réponses des diplômés des établissements ayant participé à l'enquête. De cette façon, les estimations pondérées de l'ETD représentent l'ensemble de la population, répondants et non-répondants confondus. Les facteurs de pondération ont été établis de la façon suivante :

1. Au cours de la période de collecte des données, on a effectué un suivi permanent du taux de réponse des sous-groupes de la population afin de dégager des tendances
2. On a fortement encouragé les universités à fournir d'autres renseignements sur les caractéristiques des diplômés à la fin du cycle de l'enquête afin de constituer la base définitive de sondage (voir section 6.1 ci-dessus). On disposait de ces renseignements supplémentaires sur tous les diplômés, qu'ils aient répondu ou non au questionnaire.
3. Ces renseignements additionnels fournis par les universités ont été appariés avec les répondants (et donc les non-répondants) afin de déterminer les facteurs qui pouvaient avoir un impact sur la réponse. Les femmes, par exemple, peuvent être plus portées à répondre que les hommes, et les diplômés restant au Canada plus susceptibles de répondre que ceux qui quittent le pays après l'obtention du diplôme.
4. Ces facteurs supplémentaires ont permis de regrouper les diplômés dans des groupes d'étudiants similaires (répondants comme non-répondants). On désigne également ces groupes sous le nom de catégories de poids. Chaque étudiant est placé dans une seule de ces catégories.
5. Un poids (DWEIGHT) a été calculé pour chaque répondant, égal au nombre total des diplômés dans une catégorie de poids particulière divisé par le nombre total des répondants de cette catégorie. Tous les répondants de la même catégorie se voient attribuer le même poids.

L'efficacité de cette procédure de pondération dépend en très grande partie de la disponibilité des renseignements supplémentaires sur les caractéristiques des diplômés fournis par les universités. Voir section 8.2.4 pour plus de renseignements.

8. Qualité des données

8.1 Taux de réponse

Au total, on a demandé à 3 327 diplômés de remplir le questionnaire. Cinquante-cinq d'entre eux ont refusé; 13 questionnaires ont été rejetés parce qu'ils ne comportaient pas suffisamment de réponses pour être utilisables, et 1 278 répondants n'ont pu être contactés, ce qui donne un taux de réponse de $1981/3327 = 60\%$.

8.2 Erreurs relatives à l'enquête

Des erreurs sont susceptibles de se produire à presque toutes les étapes des opérations d'enquête. Les intervieweurs ou les intervieweuses peuvent avoir mal compris les instructions, les enquêtés se tromper en répondant aux questions, les réponses être saisies de façon incorrecte sur les questionnaires et des erreurs se glisser au moment du traitement et de la totalisation des données. Il s'agit là d'autant d'erreurs non dues à l'échantillonnage.

Lorsque le nombre d'observations est élevé, les erreurs aléatoires ont peu d'effet sur les estimations calculées à partir des résultats de l'enquête. Toutefois, les erreurs systématiques contribuent à biaiser les estimations. On consacre énormément de temps et d'efforts à réduire les erreurs non dues à l'échantillonnage dans l'enquête. À chacune des étapes du cycle de collecte et de traitement des données, on applique des mesures d'assurance de la qualité pour contrôler la qualité des données. Au nombre de ces mesures figurent le recours à du personnel hautement qualifié, une formation poussée concernant les procédures et le questionnaire de l'enquête, l'observation du personnel en vue de cerner les problèmes liés à la conception du questionnaire ou à une mauvaise compréhension des instructions, des contrôles visant à réduire au minimum les erreurs de saisie des données ainsi que des vérifications du codage et des contrôles ayant pour but d'attester de la logique du traitement.

8.2.1 Base de sondage

La base de sondage de la population cible de l'ETD a été créée à partir d'une liste des établissements postsecondaires canadiens décernant un diplôme de doctorat. Cette liste est compilée et tenue à jour par le Centre de la statistique de l'éducation de Statistique Canada. Tous les établissements figurant sur la liste étaient invités à participer à l'enquête. Ceux qui n'ont pas décerné de diplôme de doctorat pendant l'année de référence étaient exclus de la population cible. On a également demandé aux établissements participant à l'enquête de fournir à Statistique Canada une liste mensuelle des nouveaux diplômés ainsi qu'une liste finale. C'est cette dernière qui servait de base de sondage.

L'Enquête auprès des titulaires d'un doctorat visait tous les établissements décernant un doctorat, mais neuf d'entre eux soit n'ont pu être contactés, ou ont été contactés mais n'ont pas participé à l'enquête, ou encore ont été contactés et voulaient participer mais ont eu des difficultés d'acheminement lors de la distribution du questionnaire. On évalue à quelque 9 %, soit environ 300 doctorants, le nombre d'étudiants qui ont obtenu un doctorat auprès de ces établissements. Aucune correction n'a été faite pour en tenir compte, la majorité d'entre eux provenant du même établissement. La plupart des caractéristiques transversales ne devraient pas être affectées, mais il ne sera pas possible de faire une comparaison directe des niveaux publiés d'une année à l'autre.

8.2.2 Collecte des données

On a remis à tous les établissements participants un Guide des personnes-ressources de l'établissement de l'ETD, ainsi que le matériel et les logiciels nécessaires. Le soutien était assuré par courrier électronique ou grâce à une ligne téléphonique sans frais. Les coordonnées pour joindre Statistique Canada étaient également fournies dans le questionnaire afin que les répondants puissent communiquer directement avec Statistique Canada s'ils avaient des questions à poser. Le personnel opérationnel de Statistique Canada affecté à l'enquête avait reçu de la formation, dont une description du contexte et des objectifs de l'enquête, ainsi qu'un glossaire des termes et de leurs définitions.

La période de collecte des données allait de septembre 2003 à septembre 2004. Les pratiques en vigueur variaient d'un établissement à l'autre, comme on le précise à la section 6.1. On ne sait pas si ces différentes pratiques avaient une incidence quelconque; cependant, on pense que le moment où le matériel d'enquête a été distribué par les établissements a eu un effet sur le taux de réponse. Par exemple, deux étudiants ont pu terminer en même temps le travail demandé pour leur doctorat, mais l'un des établissements a distribué le matériel au moment de la soutenance de thèse, tandis que l'autre l'a fait au moment de la confirmation du diplôme par les instances dirigeantes. Dans ce scénario, on aurait posté le questionnaire au deuxième étudiant bien des mois après que le premier l'ait reçu. Compte tenu de la mobilité des diplômés, la proportion des étudiants auprès de qui il n'a pas été possible de faire de suivi serait plus élevée dans le deuxième établissement que dans le premier.

8.2.3 Traitement des données

L'Enquête auprès des titulaires d'un doctorat est unique en ce sens que les fichiers constituant la base de sondage, soit les listes définitives fournies par les établissements d'enseignement, ont été reçus à la fin de la période de collecte des données, et non avant comme c'est l'habitude. Ainsi les questionnaires ont-ils donc pu se trouver remplis par des personnes qui ne se trouvaient pas sur les listes définitives, les listes mensuelles, par ailleurs, ne correspondant pas forcément exactement aux listes finales.

Le diplôme de doctorat étant décerné en tout temps, il a pu arriver que certains répondants ayant obtenu leur diplôme au début de la période de référence aient reçu le questionnaire de l'année précédente, tandis que d'autres, qui ont été diplômés à la fin de la période, peuvent fort bien avoir reçu le questionnaire du cycle suivant de l'enquête. On a corrigé ces erreurs manuellement.

8.2.4 Non-réponse

L'une des principales sources d'erreurs non dues à l'échantillonnage observées dans le cadre des enquêtes est l'effet de la non-réponse sur les résultats. L'étendue d'une non-réponse varie d'une non-réponse partielle (le fait de ne pas répondre simplement à une ou à des questions) à une non-réponse totale. S'il y a eu non-réponse totale, c'est que l'on a été incapable de communiquer avec le répondant, ou qu'il a refusé de participer à l'enquête, ou bien que le questionnaire n'était pas suffisamment complet. Dans le cas de l'ETD de 2003-2004, on a

enregistré 13 dossiers de non-réponse partielle comme étant des non-réponses totales parce qu'ils ne contenaient pas suffisamment de données.

Non-réponse totale

On a traité les non-réponses en assignant un facteur de pondération aux répondants (voir section 7.5 pour de plus amples renseignements). Le taux global de non-réponse de 40 % à l'ETD réduit la fiabilité des estimations dérivées des données de l'enquête. La présence de non-réponses lors d'un recensement entraîne une variance et éventuellement un biais des caractéristiques estimées. Le niveau de l'impact sur l'estimation d'une caractéristique dépend de la similitude entre les répondants et les non-répondants pour cette caractéristique, et dans quelle mesure les facteurs de pondération tiennent compte des dissimilarités. La méthode d'estimation de l'ETD part de l'hypothèse que toutes les personnes entrant dans une même catégorie de poids, aussi bien les répondants que les non-répondants, ont la même propension à répondre et que cette propension est indépendante des caractéristiques mesurées par l'enquête. La validité de ces hypothèses détermine la qualité des estimations dérivées de l'enquête et peut varier d'une caractéristique à l'autre.

Pour ce qui est de l'ETD de 2003-2004, l'information secondaire sur la base de sondage utilisable pour établir des catégories de poids était très limitée. Il n'a donc pas été possible de construire des catégories permettant de tenir compte de toutes les sources prévisibles de biais pour cause de non-réponse. En particulier, les estimations de l'erreur ne tiennent pas compte du biais potentiel dû à la faible proportion de réponses des titulaires qui ont quitté le Canada. Les utilisateurs de ces données sont invités à la plus grande prudence au moment d'extrapoler les résultats de l'ETD de 2003-2004 à la population des diplômés qui ont quitté le Canada immédiatement après avoir obtenu leur diplôme.

Non-réponse partielle

Dans la plupart des cas, il y a non-réponse partielle à l'enquête lorsque le répondant n'a pas compris ou a mal compris une question, qu'il a refusé de répondre à une question ou qu'il ne souvenait pas de l'information demandée. Lors du traitement des données, on a généré plus de réponses inconnues suite à l'incohérence des données ou, le plus souvent, parce qu'un volet du questionnaire avait été omis lors de la collecte. Après le traitement des données, on a classé les réponses inconnues dans la catégorie « Non déclaré », sans faire d'imputation.

Un certain nombre de questions étaient considérées comme étant cruciales pour l'analyse et on les a utilisées pour affecter un code « partiel » ou « complet » au questionnaire. Le tableau ci-dessous donne le taux de non-réponse (Non déclaré) pour ces catégories.

Taux de non-réponse aux questions essentielles

Question	Description	Taux (%)
A6	Source primaire de soutien financier pendant les études supérieures	2,1
A8 Diplôme de doctorat	Mois du début des études de doctorat	0,5
	Année du début des études de doctorat	0,6
	Mois de l'obtention du diplôme de doctorat	0,4
	Année de l'obtention du diplôme de doctorat	0,4
	Domaine des études de doctorat	0,1
A15	Endettement directement lié aux études de premier cycle	1,3
	Endettement lié aux études supérieures	1,0
B1	Pays visé après avoir obtenu le diplôme	1,6
B3	Projets à la fin des études supérieures (l'année suivante)	1,5
B4	Description des projets après l'obtention du diplôme	2,4
C1	Sexe du répondant	0,3
C7	Année de naissance	0,7
	Mois de naissance	0,9

Les autres variables du fichier de microdonnées présentent des taux similaires, à l'exception de la question B14. Dans ce cas, le taux est de près de 10 % et on recommande de faire une analyse soigneuse de cette variable.

9. Lignes directrices pour la totalisation, l'analyse et la diffusion des données

Ce chapitre donne un aperçu des lignes directrices que doivent respecter les utilisateurs qui totalisent, analysent, publient ou diffusent autrement des données calculées à partir des fichiers de microdonnées de l'enquête. Ces lignes directrices devraient permettre aux utilisateurs de microdonnées de produire les mêmes chiffres que ceux de Statistique Canada ainsi que des chiffres actuellement inédits de façon conforme à ces mêmes lignes directrices.

9.1 Lignes directrices sur l'arrondissement

Afin que les estimations qu'on doit publier ou autrement diffuser, calculées à partir de ces fichiers de microdonnées, correspondent à celles produites par Statistique Canada, nous conseillons vivement aux utilisateurs de respecter les lignes directrices qui suivent en ce qui concerne l'arrondissement :

- a) Les estimations dans le corps d'un tableau statistique doivent être arrondies à la centaine près à l'aide de la technique d'arrondissement normale. Suivant cette technique, si le premier ou le seul chiffre à supprimer se situe entre 0 et 4, on ne modifie pas le dernier chiffre à conserver. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on accroît de 1 le dernier chiffre à conserver. Par exemple, suivant la technique d'arrondissement normale à la centaine près, si les deux derniers chiffres se situent entre 00 et 49, on les remplace par 00 et on laisse inchangé le chiffre précédent (celui des centaines). Si les derniers chiffres se situent entre 50 et 99, on les remplace par 00 et on accroît de 1 le chiffre précédent.
- b) Les totaux partiels marginaux et les totaux marginaux des tableaux statistiques doivent être calculés à partir de leurs composantes non arrondies correspondantes, puis ensuite arrondis à leur tour à la centaine près à l'aide de la technique d'arrondissement normale.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de composantes non arrondies (c'est-à-dire des numérateurs et/ou des dénominateurs), puis arrondis à leur tour à une décimale à l'aide de la technique d'arrondissement normale. Dans le cas d'un arrondissement normal à un seul chiffre, si le dernier ou le seul chiffre à supprimer se situe entre 0 et 4, on ne modifie pas le dernier chiffre à conserver. Si le premier ou le seul chiffre à supprimer se situe entre 5 et 9, on accroît de 1 le dernier chiffre à conserver.
- d) Les sommes et les différences d'agrégats (ou des rapports) doivent être calculées à partir de leurs composantes non arrondies correspondantes, puis arrondies à leur tour à la centaine près (ou à la décimale près) à l'aide de la technique d'arrondissement normale.
- e) Dans les cas où, en raison de limites techniques ou d'autres limites, une technique d'arrondissement autre que la technique normale est utilisée, produisant des estimations qui doivent être publiées ou autrement diffusées, différentes des estimations correspondantes publiées par Statistique Canada, nous conseillons vivement aux utilisateurs d'indiquer la raison de ces différences dans le ou les documents qui seront publiés ou diffusés.
- f) En aucun cas, les utilisateurs ne doivent publier ou autrement diffuser des estimations non arrondies. Des estimations non arrondies laissent entendre qu'elles sont plus précises qu'elles le sont en réalité.

9.2 Lignes directrices pour la pondération d'une totalisation

Bien que l'enquête auprès des titulaires d'un doctorat (ETD) soit conçue en tant que recensement, des poids ont été appliqués. Les utilisateurs devraient donc utiliser le facteur de pondération approprié lorsqu'ils produisent des estimations.

Si l'on n'utilise pas des poids appropriés, on ne peut considérer les estimations calculées à partir des fichiers de microdonnées comme représentatives de la population visée par l'enquête et ces estimations ne correspondront pas à celles produites par Statistique Canada.

Les utilisateurs devraient également noter que certains progiciels pourraient ne pas permettre la production d'estimations correspondant exactement à celles qu'offre Statistique Canada, en raison du mode de traitement du champ du poids.

9.3 Définitions des types d'estimations : catégoriques et quantitatives

Avant de discuter de la façon dont on peut totaliser et analyser les données de l'ETD, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques de la population que l'on peut produire à partir du fichier de microdonnées créé pour l'ETD.

9.3.1 Estimations catégoriques

Les estimations catégoriques sont des estimations du nombre ou du pourcentage de membres de la population visée par l'enquête, possédant certaines caractéristiques ou faisant partie d'une catégorie définie. Le nombre de diplômés en sciences biologiques titulaires d'un visa ou leur proportion constituent des exemples de telles estimations. On peut aussi appeler une estimation du nombre de personnes possédant une certaine caractéristique une estimation d'un agrégat.

Exemples de questions catégoriques :

Q: Dans quel pays comptez-vous vivre après avoir obtenu votre diplôme (d'ici un an)?

A: Au Canada / Aux États-Unis / Autre pays / Non déclaré

Q: Comptez-vous faire un stage postdoctoral?

A: Oui / Non / Non déclaré

9.3.2 Estimations quantitatives

Les estimations quantitatives sont des estimations de totaux ou de moyennes, de médianes ou d'autres mesures d'une tendance centrale de quantités, reposant sur quelques-uns ou sur la totalité des membres de la population visée par l'enquête.

Elles comprennent aussi expressément des estimations de la forme \hat{X} / \hat{Y} , où \hat{X} est une estimation de la quantité totale de membres de la population visée par l'enquête et \hat{Y} , une estimation du nombre de personnes de la population visée par l'enquête ayant contribué à en arriver à cette quantité totale.

Le temps moyen pour terminer des études doctorales par les individus qui ont reçu un prêt d'une source quelconque constitue un exemple d'estimation quantitative. Le numérateur est une estimation du temps total consacré par les étudiants qui ont reçu

un prêt d'une source quelconque, et son dénominateur est une estimation du nombre de diplômés dans ce cas.

Exemple de question quantitative :

Q: Une fois les cours et les examens terminés, pendant combien de mois ou d'années avez-vous travaillé à votre thèse (préparation ou recherche non liée à un cours, rédaction et soutenance)?

A: Mois |_|_| OU Années |_|_|

9.3.3 Totalisation d'estimations catégoriques

On peut obtenir des estimations du nombre de gens possédant une certaine caractéristique à partir du fichier de microdonnées en additionnant les poids finals de tous les enregistrements présentant la ou les caractéristiques qui nous intéressent.

On obtient des proportions et des rapports de la forme \hat{X}/\hat{Y} en :

- additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le numérateur (\hat{X}) ,
- additionnant les poids finals des enregistrements présentant la caractéristique qui nous intéresse pour le dénominateur (\hat{Y}) , puis en
- divisant l'estimation a) par celle de b) (\hat{X}/\hat{Y}) .

9.3.4 Totalisation d'estimations quantitatives

On peut obtenir des estimations de quantités à partir du fichier de microdonnées en multipliant la valeur de la variable qui nous intéresse par le poids final établi pour chaque enregistrement, puis en additionnant cette quantité pour tous les enregistrements qui nous intéressent. Pour obtenir une moyenne pondérée de la forme \hat{X}/\hat{Y} , on calcule le numérateur (\hat{X}) comme dans le cas d'une estimation quantitative et on calcule le dénominateur (\hat{Y}) comme dans celui d'une estimation catégorique. Par exemple, on peut produire une estimation du nombre moyen de mois nécessaires pour terminer un doctorat en génie en procédant comme suit :

- Tenir compte uniquement des enregistrements où la valeur est consignée, soit lorsque le nombre de mois, DV_TIMPH, n'est pas égal à 999 et que PHDNUMB se situe entre 300 et 399 (génie).
- Pour chaque enregistrement, multiplier la valeur indiquée dans la variable dérivée DV_TIMPH par DWEIGHT, le poids définitif de l'enregistrement.
- Additionner les produits calculés en b).
- Additionner les valeurs de DWEIGHT pour tous les enregistrements où PHDNUMB se situe entre 300 et 399 (génie).
- Diviser la somme de c) par la somme de d).

9.4 Lignes directrices pour l'analyse statistique

Bien que l'ETD s'apparente à un recensement, les non-réponses constituent une source significative d'erreur dans les estimations produites à partir de l'enquête. On a recours à un ajustement des données au moyen de poids pour compenser les non-réponses, et à un estimateur de variance pour faire l'estimation des erreurs dues aux non-réponses. Il faut employer les poids de l'enquête pour que les estimations et les analyses de données d'enquête soient le plus exemptes de biais possible.

Même si beaucoup de procédures d'analyse qu'on trouve à l'intérieur de progiciels statistiques permettent l'utilisation de poids, la signification ou la définition du poids inclus dans ces procédures peut différer de celle qui convient dans une base de sondage, de telle sorte que bien que souvent les estimations produites par les progiciels soient correctes, les variances calculées sont elles mauvaises. Les variances approximatives pour des estimations simples comme des totaux, des proportions et des rapports (pour des variables qualitatives) peuvent être dérivées à partir du tableau de variabilité approximative inclus à la section 10.5.

Pour d'autres techniques d'analyse (de régression linéaire, de régression logistique et de l'analyse de variance, par exemple), il existe une méthode qui peut rendre les variances calculées par l'application des progiciels normalisés plus significatives, en intégrant les probabilités inégales de sélection. L'application de cette méthode entraîne une remise à l'échelle des poids qui fait que le poids moyen est de 1 et la somme des poids égale au nombre de répondants.

Supposons, par exemple, qu'il faille analyser tous les cas de répondants de sexe masculin. Les étapes à suivre pour remettre à l'échelle les poids sont les suivantes :

- 1) sélectionner tous les répondants du fichier qui ont déclaré SEXE = homme;
- 2) calculer le poids MOYEN pour ces enregistrements en additionnant DWEIGHT, les poids originaux des gens établis à partir du fichier de microdonnées pour ces enregistrements, puis le diviser par le nombre de répondants ayant déclaré SEXE = homme;
- 3) pour chacun de ces répondants, calculer un poids REMIS À L'ÉCHELLE égal au poids original des gens divisé par le poids MOYEN;
- 4) effectuer l'analyse pour ces répondants à l'aide du poids REMIS À L'ÉCHELLE.

Pour calculer une estimation plus précise de la variance, il faudrait disposer d'autres renseignements. Ces chiffres sont présentés dans un fichier distinct, que les utilisateurs peuvent consulter s'ils ont accès à un logiciel qui peut interpréter un plan d'échantillonnage aléatoire simple. L'ETD n'était pas une enquête-échantillon, mais les renseignements descriptifs et la variance permettent d'estimer l'erreur due aux non-réponses. Pour obtenir ces estimations de l'erreur, le logiciel doit offrir l'option d'échantillonnage aléatoire simple stratifié, les catégories de poids de l'ETD pouvant être considérées comme des strates. Lorsqu'on utilise un tel logiciel, DWEIGHT, plutôt que le poids remis à l'échelle décrit plus haut, est la variable de poids appropriée. Les utilisateurs peuvent communiquer avec Statistique Canada pour tout conseil concernant la formule adéquate à utiliser pour faire ces calculs.

9.5 Lignes directrices sur la diffusion de coefficients de variation

Avant de diffuser et/ou de publier toute estimation établie à partir de l'ETD, les utilisateurs devraient premièrement déterminer le niveau de qualité de cette estimation. Les niveaux de qualité sont *acceptable*, *médiocre* et *inacceptable*. Les erreurs non dues à l'échantillonnage influencent la qualité des données comme il en a été question au chapitre 8. Aux fins du présent document, cependant, on ne déterminera le niveau de qualité d'une estimation qu'à partir de l'estimation d'une erreur due à la non-réponse dont rend compte le coefficient de variation indiqué à l'intérieur du tableau qui figure ci-dessous. Les utilisateurs devraient néanmoins s'assurer de lire le chapitre 8 pour être plus pleinement informés des caractéristiques relatives à la qualité de ces données.

On devrait premièrement déterminer le nombre de répondants retenus pour le calcul de l'estimation. Si ce nombre est inférieur à cinq, il faudrait considérer l'estimation pondérée de qualité inacceptable.

Pour les estimations pondérées reposant sur les tailles d'échantillons de cinq répondants ou plus, les utilisateurs devraient déterminer le coefficient de variation de l'estimation et suivre les lignes directrices qui figurent ci-dessous. On devrait appliquer ces lignes directrices relatives au niveau de qualité aux estimations arrondies pondérées

On peut considérer qu'il est possible de divulguer toutes les estimations. Celles d'un niveau de qualité médiocre ou inacceptable doivent cependant être accompagnées d'une mise en garde pour avertir les utilisateurs subséquents.

Lignes directrices relatives au niveau de qualité de l'estimation

Niveau de qualité de l'estimation	Lignes directrices
1) Acceptable	<p>Estimations : d'une taille d'échantillon de 5 ou plus et des coefficients de variation de l'ordre de 0,0 à 16,5 %.</p> <p>Aucune mise en garde n'est nécessaire.</p>
2) Médiocre	<p>Estimations : d'une taille d'échantillon de 5 ou plus et de coefficients de variation de l'ordre de 16,6 à 33,3 %.</p> <p>Ces estimations devraient être signalées par la lettre M (ou un quelconque identificateur similaire). Elles devraient être accompagnées d'une mise en garde avertissant les utilisateurs subséquents des niveaux élevés d'erreur qui y sont reliés.</p>
3) Inacceptable	<p>Estimations : d'une taille d'échantillon de moins de 5 ou de coefficients de variation de plus de 33,3 %.</p> <p>Statistique Canada recommande de ne pas diffuser d'estimations de qualité inacceptable. Si un utilisateur choisit cependant de le faire, ces estimations devraient alors être signalées à l'aide de la lettre I (ou d'un quelconque identificateur similaire) et accompagnées de la mise en garde suivante :</p> <p>« Nous informons l'utilisateur que ces estimations (désignées avec la lettre I) ne respectent pas les normes de qualité de Statistique Canada. Les conclusions qui reposeront sur ces données ne seront ni fiables ni valables très probablement. »</p>

9.6 Seuils pour la diffusion des estimations pour l'Enquête de 2003-2004 auprès des titulaires d'un doctorat

Le tableau ci-dessous fournit une indication de la précision des estimations de population ainsi que les seuils de diffusion associés aux trois niveaux de qualité de l'estimation présentés à la section précédente. Ces seuils proviennent du tableau de coefficients de variation (CV) dont il sera question au chapitre 10.

Par exemple, d'après le tableau, le coefficient de variation d'une estimation pondérée de 25 diplômés en génie se situe entre 16,6 % et 33,3 %. Ce CV est d'une qualité médiocre.

Veillez noter que ces seuils de diffusion ne correspondent qu'aux estimations des totaux de population. Dans le cas d'estimations de rapports, les utilisateurs ne devraient pas utiliser la valeur du numérateur (ni du dénominateur) afin de trouver le niveau de qualité de l'estimation correspondant. La règle 4 de la section 10.1 ainsi que l'exemple 4 de la section 10.1.1 expliquent la bonne procédure à suivre dans le cas d'un rapport.

Domaine d'études	CV acceptable 0,0 % – 16,5 %	CV médiocre 16,6 % – 33,3 %	CV inacceptable > 33,3 %
Sciences agricoles	30 et plus	8 à < 30	inf. à 8
Sciences biologiques	30 et plus	8 à < 30	inf. à 8
Sciences de la santé	25 et plus	7 à < 25	inf. à 7
Génie	30 et plus	9 à < 30	inf. à 9
Informatique et science de l'information	25 et plus	11 à < 25	inf. à 11
Mathématiques	35 et plus	11 à < 35	inf. à 11
Astronomie	Aucun	6 à < 15	inf. à 6
Science de l'atmosphère et météorologie	Aucun	9 à < 11	inf. à 9
Chimie	30 et plus	9 à < 30	inf. à 9
Géologie et sciences connexes	30 et plus	9 à < 30	inf. à 9
Physique	25 et plus	9 à < 25	inf. à 9
Sciences physiques diverses	30 et plus	9 à < 30	inf. à 9
Psychologie	25 et plus	9 à < 25	inf. à 9
Sciences sociales	30 et plus	9 à < 30	inf. à 9
Sciences humaines	25 et plus	7 à < 25	inf. à 7
Enseignement	30 et plus	6 à < 30	inf. à 6
Domaines professionnels	30 et plus	9 à < 30	inf. à 9
Non déclaré	Aucun	Aucun	Tous
Tous les diplômés	15 et plus	8 à < 15	inf. à 8

10 Tableau de variabilité approximative

Afin de fournir des coefficients de variation (CV) qui pourraient s'appliquer à un grand nombre d'estimations catégoriques produites à partir de ce fichier de microdonnées et auxquels il serait facilement possible pour l'utilisateur d'avoir accès, un tableau de variabilité approximative a été produit. Ce tableau de CV permet à l'utilisateur d'obtenir un coefficient approximatif de variation fondé sur la taille de l'estimation calculée à partir des données d'enquête.

L'ETD n'étant pas une enquête-échantillon, les estimations des CV ne constituent pas une estimation de la variabilité de l'échantillon. Dans l'ETD, le CV permet d'estimer l'erreur due aux non-réponses. Cependant, les coefficients de variation sont dérivés d'une formule de la variance pour un échantillonnage aléatoire simple stratifié. Les coefficients de variation sont calculés pour toute une série de caractéristiques, puis on retient parmi eux une valeur prudente (le 75^e centile) que l'on utilise dans le tableau de CV et qui s'applique ensuite à l'ensemble des caractéristiques.

On peut utiliser le tableau de variabilité approximative pour tous les diplômés, ainsi que pour les diplômés de presque tous les grands domaines d'études. Il y avait trop peu de diplômés en astronomie ou en science de l'atmosphère et météorologie pour produire des estimations avec un CV acceptable.

Tous les coefficients de variation inclus dans le tableau de variabilité approximative sont approximatifs et donc non officiels. On peut obtenir de Statistique Canada, contre remboursement des frais, des estimations de la variance plus précises pour des estimations ponctuelles spécifiques ou les calculer avec le logiciel approprié. Étant donné que le CV approximatif est une estimation prudente, l'utilisation d'une estimation de la variance calculée pour une estimation ponctuelle spécifique peut provoquer le changement de la qualité de l'estimation d'un niveau de qualité à un autre. Par exemple, une estimation *médiocre* pourrait devenir *acceptable* si elle était fondée sur le calcul du CV exact, l'inverse pouvant également se produire. Les utilisateurs peuvent communiquer avec Statistique Canada afin d'en obtenir des conseils sur les formules appropriées à employer pour ces calculs.

Rappel : Si le nombre d'observations sur lesquelles une estimation est basée est inférieur à 5, l'estimation pondérée est très probablement considérée de qualité inacceptable et Statistique Canada recommande de ne pas diffuser une telle estimation, indépendamment de la valeur du coefficient de variation.

10.1 Comment utiliser le tableau de coefficients de variation pour des estimations catégoriques

Les règles qui suivent devraient permettre à l'utilisateur de déterminer les coefficients approximatifs de variation à partir du tableau de variabilité approximative pour des estimations du nombre, de la proportion ou du pourcentage de membres de la population visée par l'enquête possédant une certaine caractéristique et pour des rapports et des différences entre de telles estimations.

Règle 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Le coefficient de variation dépend uniquement de la taille de l'estimation elle-même. Dans le tableau de variabilité approximative, repérez le nombre estimé dans la colonne la plus à gauche (intitulée « Numérateur du pourcentage ») et suivez les astérisques (le cas échéant) jusqu'au premier chiffre rencontré. Ce chiffre est le coefficient de variation approximatif.

Règle 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Le coefficient de variation d'une proportion estimée ou d'un pourcentage estimé dépend à la fois de la taille de la proportion ou du pourcentage et de la taille du total sur lequel la proportion ou le pourcentage repose. Les proportions estimées ou les pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, lorsque la proportion ou le pourcentage repose sur un sous-groupe de la population. La proportion, par exemple, des diplômés en sciences de la vie qui sont mariés est plus fiable que le nombre estimé de ces mêmes diplômés. (Remarquez que, dans le tableau, la valeur des coefficients de variation diminue lorsqu'on les lit de gauche à droite.)

Lorsque la proportion ou le pourcentage repose sur la population totale des diplômés, le CV de la proportion ou du pourcentage est le même que celui du numérateur de la proportion ou du pourcentage. On peut, dans ce cas, appliquer la règle 1.

Lorsque la proportion ou le pourcentage repose sur un sous-ensemble de la population totale (p. ex., les étudiants d'un domaine d'études particulier), on devrait faire référence à la proportion ou au pourcentage (dans le haut du tableau) et au numérateur de la proportion ou du pourcentage (dans la colonne de gauche du tableau). L'intersection de la rangée et de la colonne appropriées donne le coefficient de variation.

Règle 3 : Estimations de différences entre des agrégats ou des pourcentages

L'erreur-type d'une différence entre deux estimations est approximativement égale à la racine carrée de la somme des carrés de chaque erreur-type considérée séparément. C'est-à-dire que l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1, \hat{X}_2 est l'estimation 2, et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement. Le CV de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d}$. Cette formule est exacte pour la différence entre des caractéristiques distinctes et non corrélées, mais n'est autrement qu'approximative.

Règle 4 : Estimations de rapports

Si le numérateur est un sous-ensemble du dénominateur, on devrait convertir le rapport en un pourcentage et utiliser la règle 2. Cela s'appliquerait, par exemple, si le dénominateur était le nombre de diplômés en sciences de la vie et le numérateur, celui des diplômés en sciences de la vie qui sont mariés.

Si le numérateur n'est pas un sous-ensemble du dénominateur, comme dans l'exemple du rapport du nombre de femmes diplômées en enseignement comparativement au nombre de diplômés qui le sont, l'écart-type du rapport des estimations est approximativement égal à la racine carrée de la somme des carrés de chaque coefficient de variation considéré séparément, multipliée par \hat{R} . C'est-à-dire que l'erreur-type d'un rapport ($\hat{R} = \hat{X}_1 / \hat{X}_2$) est :

$$\sigma_{\hat{R}} = \hat{R} \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement. Le

coefficient de variation de \hat{R} est donné par $\sigma_{\hat{R}} / \hat{R}$. La formule tendra à surestimer l'erreur si \hat{X}_1 et \hat{X}_2 sont corrélés positivement, et à la sous-estimer si \hat{X}_1 et \hat{X}_2 le sont négativement.

Règle 5 : Estimations de différences de rapports

Dans ce cas, on combine les règles 3 et 4. On détermine premièrement les CV pour les deux rapports à l'aide de la règle 4, puis on trouve le CV de leur différence au moyen de la règle 3.

10.1.1 Exemples d'utilisation du tableau de coefficients de variation pour des estimations catégoriques

Exemple 1 : Estimations du nombre de personnes possédant une caractéristique donnée (agrégats)

Supposons qu'un utilisateur estime à 1520 le nombre d'étudiants qui n'avaient aucune dette au moment de l'obtention de leur diplôme. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Se reporter au tableau des coefficients de variation.
- 2) L'agrégat estimé (1520) ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 1500.
- 3) On trouve le coefficient de variation pour un agrégat estimé en se reportant à la première entrée autre que des astérisques sur cette rangée, c'est-à-dire 1,6 %.
- 4) Le coefficient de variation approximatif de l'estimation est donc 1,6 %. Le résultat selon lequel il y avait 1520 étudiants (à arrondir d'après les lignes directrices tel qu'il est indiqué à la section 9.1) qui n'avaient aucune dette au moment de l'obtention de leur diplôme peut être publié sans réserve.

Exemple 2 : Estimations de proportions ou de pourcentages de personnes possédant une caractéristique donnée

Supposons qu'un utilisateur estime à $402 / 945 = 42.5\%$ la proportion de diplômés en sciences de la vie qui n'avaient aucune dette au moment de l'obtention de leur diplôme. Comment l'utilisateur détermine-t-il le coefficient de variation de cette estimation?

- 1) Se reporter au tableau des coefficients de variation.
- 2) Parce que l'estimation est un pourcentage fondé sur un sous-ensemble de la population totale (c'est-à-dire les diplômés en sciences de la vie qui n'avaient aucune dette au moment de l'obtention de leur diplôme), il faut utiliser à la fois le pourcentage (42,5 %) et la portion numérateur du pourcentage (402) pour déterminer le coefficient de variation.
- 3) Le numérateur, 402, ne figure pas dans la colonne de gauche (la colonne « Numérateur du pourcentage »); il faut donc utiliser le chiffre qui s'en rapproche le plus, c'est-à-dire 400.

- 4) Le pourcentage estimé, 42,5 %, se situant à mi-chemin entre 35 % et 50 %, il est donc nécessaire de retenir le chiffre le plus bas, 35 %.
- 5) Le chiffre indiqué à l'intersection de la rangée et de la colonne utilisées, c'est-à-dire 4,1 %, est le coefficient de variation à employer.

Le coefficient de variation approximatif de l'estimation est donc 4,1 %. Le résultat selon lequel 42,5 % des diplômés en sciences de la vie n'avaient aucune dette au moment de l'obtention de leur diplôme peut être publié sans réserve.

Exemple 3 : Estimations de différences entre des agrégats ou des pourcentages

Supposons qu'un utilisateur estime à $402 / 945 = 42,5 \%$ la proportion de diplômés en sciences de la vie qui n'avaient pas de dettes au moment de l'obtention de leur diplôme par rapport à $230 / 599 = 38,4 \%$ celle des diplômés en sciences sociales dans le même cas. Comment l'utilisateur détermine-t-il le coefficient de variation de la différence entre ces deux estimations?

- 1) L'utilisation du tableau des coefficients de variation de la même manière que celle décrite dans l'exemple 2 donne un CV de l'estimation pour les diplômés en sciences de la vie qui n'avaient pas de dettes au moment de l'obtention de leur diplôme de 4,1 %.

L'utilisation du tableau des coefficients de variation de la même manière que celle décrite dans l'exemple 2 donne un CV de l'estimation pour les diplômés en sciences sociales qui n'avaient pas de dettes au moment de l'obtention de leur diplôme de 5,5 %.

- 2) Si l'on utilise la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{X}_1 - \hat{X}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{X}_1 \alpha_1)^2 + (\hat{X}_2 \alpha_2)^2}$$

où \hat{X}_1 est l'estimation 1 (sciences de la vie), \hat{X}_2 est l'estimation 2 (sciences sociales), et α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et \hat{X}_2 respectivement.

Autrement dit, l'erreur-type de la différence $\hat{d} = 0,425 - 0,384 = 0,041$ est :

$$\begin{aligned} \sigma_{\hat{d}} &= \sqrt{[(0,425)(0,0041)]^2 + [(0,384)(0,055)]^2} \\ &= \sqrt{(0,0003036) + (0,000446054)} \\ &= 0,027 \end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,027 / 0,041 = 0,658$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 65.8%. Le résultat selon lequel la proportion d'étudiants n'ayant pas de dettes au moment de l'obtention de leur diplôme est plus élevée de 4,1 % dans le cas des sciences de la vie que dans celui des sciences sociales est considéré de qualité inacceptable et Statistique Canada recommande de ne pas le publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra

être désignée ainsi en utilisant la lettre I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les prochains utilisateurs en garde contre les hauts taux d'erreur associés à l'estimation.

Exemple 4 : Estimations de rapports

Supposons qu'un utilisateur estime à 891 le nombre de diplômés de sexe masculin qui n'avaient pas de dettes au moment de l'obtention de leur diplôme, par rapport à 629 dans le cas des femmes. L'utilisateur est intéressé à comparer l'estimation des hommes à celle des femmes sous la forme d'un rapport. Comment détermine-t-il le CV de cette estimation?

- 1) Tout d'abord, cette estimation est une estimation d'un rapport, où le numérateur de l'estimation (\hat{X}_1) est le nombre d'hommes qui n'avaient pas de dettes au moment de l'obtention de leur diplôme. Le dénominateur de l'estimation (\hat{X}_2) est le nombre de femmes dans le même cas.
- 2) Se reporter au tableau des coefficients de variation.
- 3) Le numérateur de cette estimation de rapport est 891. Le chiffre qui s'en rapproche le plus est 1000. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette ligne, c'est-à-dire 2,3 %.
- 4) Le dénominateur de cette estimation de rapport est 629. Le chiffre qui s'en rapproche le plus est 750. On trouve le coefficient de variation pour cette estimation en se reportant à la première entrée autre que des astérisques sur cette ligne, c'est-à-dire 2,9 %.
- 5) Le coefficient approximatif de variation de l'estimation du rapport est donc donné par la règle 4, qui est :

$$\alpha_{\hat{R}} = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{X}_1 et de \hat{X}_2 respectivement.

Autrement dit :

$$\begin{aligned}\alpha_{\hat{R}} &= \sqrt{(0,023)^2 + (0,029)^2} \\ &= \sqrt{0,000529 + 0,000841} \\ &= 0,037\end{aligned}$$

Le rapport obtenu dans le cas des hommes qui n'avaient pas de dettes au moment de l'obtention de leur diplôme par rapport aux femmes dans une situation équivalente est 891 / 629, c'est-à-dire 1,42:1 (à arrondir d'après les lignes directrices tel qu'il est indiqué à la section 9.1). Le coefficient de variation de cette estimation est 3,7 %, ce qu'on peut diffuser sans réserve.

Exemple 5 : Estimations de différences de rapports

Supposons qu'un utilisateur estime que le rapport hommes/femmes des diplômés en sciences de la vie qui n'avaient pas de dettes au moment de l'obtention de leur

diplôme est 203 pour 199, ou 1,02:1, tandis qu'il est de 98 pour 132, ou 0,74:1, dans le cas des étudiants en sciences sociales dans la même situation. L'utilisateur voudra comparer les deux rapports pour voir s'il y a une différence statistique entre eux. Comment détermine-t-il le coefficient de variation de cette différence?

- 1) On calcule tout d'abord le coefficient de variation approximatif pour le rapport des sciences de la vie (\hat{R}_1) et pour celui des sciences sociales (\hat{R}_2) tel qu'il est décrit dans l'exemple 4. Le CV approximatif pour le rapport des sciences de la vie est 8,5 % et 11,9 % pour celui des sciences sociales.
- 2) Si l'on utilise la règle 3, l'erreur-type d'une différence ($\hat{d} = \hat{R}_1 - \hat{R}_2$) est :

$$\sigma_{\hat{d}} = \sqrt{(\hat{R}_1 \alpha_1)^2 + (\hat{R}_2 \alpha_2)^2}$$

où α_1 et α_2 sont les coefficients de variation de \hat{R}_1 et \hat{R}_2 respectivement.

Autrement dit, l'erreur-type de la différence $\hat{d} = 1,02 - 0,74 = 0,28$ est :

$$\begin{aligned}\sigma_{\hat{d}} &= \sqrt{[(1,02)(0,085)]^2 + [(0,74)(0,119)]^2} \\ &= \sqrt{(0,00752) + (0,00775)} \\ &= 0,124\end{aligned}$$

- 3) Le coefficient de variation de \hat{d} est donné par $\sigma_{\hat{d}} / \hat{d} = 0,124 / 0,28 = 0,443$.
- 4) Le coefficient de variation approximatif de la différence entre les estimations est donc 44,3 %. La constatation que le rapport hommes/femmes des étudiants sans dettes est supérieur de 0,28:1 dans les sciences de la vie par rapport aux sciences sociales est considéré de qualité inacceptable et Statistique Canada recommande de ne pas le publier. Cependant, si l'utilisateur choisit de publier cette donnée, elle devra être désignée ainsi en utilisant la lettre I (ou un autre identificateur semblable) et être accompagnée d'un avertissement mettant les prochains utilisateurs en garde contre les hauts taux d'erreur associés à l'estimation.

10.2 Comment utiliser le tableau des coefficients de variation pour obtenir des limites de confiance

Même si l'on utilise beaucoup les coefficients de variation, l'intervalle de confiance d'une estimation est une mesure plus intuitivement significative de l'erreur. Un intervalle de confiance constitue une déclaration au sujet du niveau de confiance, selon laquelle la valeur vraie pour la population se situe à l'intérieur d'une gamme précisée de valeurs. On peut, par exemple, décrire comme suit un intervalle de confiance de 95 % :

Si l'on répète indéfiniment l'échantillonnage de la population, et qu'elle présente les mêmes mécanismes et les mêmes schémas de réponse, chaque échantillon menant à un nouvel intervalle de confiance pour une estimation, l'intervalle englobera alors dans 95 % des échantillons la valeur vraie de la population.

On peut obtenir, à l'aide de l'erreur-type d'une estimation, des intervalles de confiance pour des estimations en partant de l'hypothèse qu'aux termes d'un échantillonnage répété de la population, les diverses estimations obtenues pour une caractéristique donnée de la population se répartiront normalement autour de la valeur vraie de la population. Selon cette hypothèse, il y a environ 68 chances sur 100 que l'écart entre une estimation de l'échantillon et la valeur vraie pour la population soit inférieur à une erreur-type, environ 95 chances sur 100 que l'écart soit inférieur à deux erreurs-types, et environ 99 chances sur 100 qu'il soit inférieur à trois erreurs-types. On appelle ces différents degrés de confiance des niveaux de confiance.

On exprime généralement des intervalles de confiance pour une estimation, \hat{X} , sous forme de deux chiffres, un inférieur et un supérieur à l'estimation, comme étant $(\hat{X} - k, \hat{X} + k)$, où on détermine k suivant le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement des intervalles de confiance pour une estimation à partir du tableau de variabilité approximative, en déterminant d'abord à partir de la table appropriée le CV de l'estimation \hat{X} , puis en utilisant la formule suivante pour le convertir à un intervalle de confiance ($IC_{\hat{x}}$) :

$$IC_{\hat{x}} = (\hat{X} - t\hat{X}\alpha_{\hat{x}}, \hat{X} + t\hat{X}\alpha_{\hat{x}})$$

où $\alpha_{\hat{x}}$ est le coefficient de variation déterminé de \hat{X} , et

- $t = 1$ si l'on désire un intervalle de confiance de 68 %;
- $t = 1,6$ si l'on désire un intervalle de confiance de 90 %;
- $t = 2$ si l'on désire un intervalle de confiance de 95 %;
- $t = 2,6$ si l'on désire un intervalle de confiance de 99 %.

Nota : Ces valeurs de t ne sont pas exactes, mais sont acceptables selon les lignes directrices sur la diffusion des résultats.

Les lignes directrices sur la diffusion des estimations s'appliquent également aux intervalles de confiance. S'il est impossible, par exemple, de diffuser une estimation, on ne peut alors pas non plus communiquer un intervalle de confiance.

10.2.1 Exemple d'utilisation du tableau de coefficients de variation pour obtenir les limites de confiance

On calculerait comme suit un intervalle de confiance de 95 % pour la proportion estimée de diplômés en sciences de la vie qui n'avaient pas de dettes au moment de l'obtention de leur diplôme (d'après l'exemple 2 de la section 10.1.1) :

$$\hat{X} = 42,5 \% \text{ (ou exprimé sous forme de proportion, } 0,425)$$

$$t = 2$$

$\alpha_{\hat{x}}$ = 4,1 % (0,041 exprimé sous forme de proportion) est le coefficient de variation de cette estimation, déterminé à partir du tableau.

$$IC_{\hat{x}} = \{0,425 - (2) (0,425) (0,041), 0,425 + (2) (0,425) (0,041)\}$$

$$IC_{\hat{x}} = \{0,425 - 0,035, 0,425 + 0,035\}$$

$$IC_{\hat{x}} = \{0,39, 0,46\}$$

Avec un intervalle de confiance de 95 %, on peut dire qu'entre 39 % et 46 % des diplômés en sciences de la vie n'avaient pas de dettes au moment de l'obtention de leur diplôme.

10.3 Comment utiliser le tableau de coefficients de variation pour effectuer un test t

On peut aussi utiliser des erreurs-types pour effectuer des tests d'hypothèses, une procédure destinée à distinguer des paramètres d'une population à l'aide d'estimations d'un échantillon. Ces estimations peuvent être des chiffres, des moyennes, des pourcentages, des rapports, 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 quand, en fait, elles sont identiques.

Supposons que \hat{X}_1 et \hat{X}_2 soient des estimations d'un échantillon pour deux caractéristiques qui nous intéressent. Supposons également que l'erreur-type de la différence $\hat{X}_1 - \hat{X}_2$ soit $\sigma_{\hat{d}}$.

$$\text{Si } t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_{\hat{d}}}$$

se situe entre -2 et 2, aucune conclusion à propos de la différence entre les caractéristiques n'est alors justifiée au niveau de signification de 5 %. Si, cependant, ce rapport est inférieur à -2 ou supérieur à +2, la différence observée est significative au niveau de 0,05. C'est-à-dire que la différence entre les estimations est significative.

10.3.1 Exemple d'utilisation du tableau de coefficients de variation pour effectuer un test t

Supposons que l'utilisateur désire tester, au niveau de signification de 5 %, l'hypothèse selon laquelle il n'y a pas de différence entre la proportion de diplômés en sciences de la vie qui n'avaient pas de dettes au moment de l'obtention de leur diplôme et celle des diplômés en sciences sociales dans le même cas. D'après l'exemple 3 de la section 10.1.1, il s'est avéré que l'erreur-type de la différence entre ces deux estimations était 0,027. Par conséquent,

$$t = \frac{\hat{X}_1 - \hat{X}_2}{\sigma_d} = \frac{0,425 - 0,384}{0,027} = \frac{-0,041}{0,027} = 1,5$$

Puisque $t = 1,5$ est situé entre -2 et 2, il faut en conclure qu'il n'existe pas de différence significative entre les deux estimations au niveau de signification de 0,05.

10.4 Coefficients de variation pour des estimations quantitatives

Il faudrait produire des tableaux spéciaux afin de déterminer l'erreur d'échantillonnage d'estimations quantitatives, ce qui n'a pas été fait, parce que la plupart des variables pour l'ETD sont principalement de nature catégorique.

En général cependant, le coefficient de variation d'un total quantitatif sera supérieur au coefficient de variation de l'estimation de la catégorie correspondante (c'est-à-dire l'estimation du nombre de personnes retenues dans l'estimation quantitative). S'il est impossible de diffuser l'estimation de la catégorie correspondante, on ne pourra pas non plus communiquer l'estimation quantitative. Par exemple, le coefficient de variation du total des prêts étudiants consentis par le gouvernement serait supérieur au coefficient de variation du nombre correspondant d'individus. S'il est impossible, par conséquent, de diffuser le nombre estimé d'individus, on ne pourra probablement pas non plus communiquer l'estimation quantitative correspondante.

On peut calculer au besoin des coefficients de variation pour des estimations ponctuelles précises. Les utilisateurs désireux de calculer des coefficients de variation pour des estimations quantitatives peuvent utiliser un logiciel statistique qui peut s'accommoder d'un plan d'échantillonnage aléatoire simple stratifié, ainsi que du fichier supplémentaire. Ils peuvent également communiquer avec Statistique Canada afin d'obtenir des conseils sur les formules à employer pour ces calculs.

10.5 Tableau de variabilité approximative

Numérateur du pourcentage	Pourcentage											
	5	10	15	25	35	50	65	75	85	90	95	
5	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0	30.0
10	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0	28.0
12	26.0	26.0	26.0	26.0	26.0	26.0	26.0	26.0	26.0	26.0	26.0	26.0
15	24.0	23.8	23.6	23.6	23.6	23.6	23.6	23.5	23.5	23.4	23.4	23.4
20	21.0	21.0	21.0	21.0	21.4	21.9	21.6	21.3	19.8	19.8	19.8	19.8
25	18.0	18.0	18.0	17.9	17.9	17.9	17.9	17.9	17.9	17.9	17.9	17.9
30	16.5	16.5	16.5	16.4	16.4	16.3	16.3	16.3	16.3	16.3	16.3	16.3
40	14.6	14.6	14.5	14.5	14.5	14.5	14.4	14.4	14.4	14.4	14.4	14.4
50	13.2	13.1	13.1	13.0	13.0	13.0	12.8	12.6	12.6	12.6	12.6	12.6
60	11.7	11.6	11.6	11.6	11.5	11.5	11.5	11.3	11.3	11.3	11.3	11.2
70	11.1	11.1	11.1	11.1	11.1	11.0	11.0	11.0	11.0	11.0	11.0	11.0
80	10.0	10.0	9.9	9.9	9.8	9.8	9.8	9.8	9.8	9.8	9.8	9.8
90	9.3	9.3	9.3	9.2	9.2	9.2	9.0	9.0	9.0	9.0	9.0	9.0
100	9.0	9.0	9.0	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9	8.9
125	7.8	7.7	7.7	7.7	7.7	7.7	7.5	7.5	7.5	7.5	7.5	7.5
150	6.9	6.9	6.9	6.9	6.9	6.8	6.8	6.8	6.8	6.8	6.8	6.8
175	6.5	6.5	6.5	6.5	6.4	6.4	6.4	6.4	6.3	6.3	6.3	6.3
200	6.0	6.0	6.0	6.0	6.0	6.0	5.9	5.9	5.8	5.8	5.8	5.8
250	*****	5.5	5.5	5.5	5.5	5.5	5.4	5.4	5.4	5.4	5.4	5.4
300	*****	4.9	4.8	4.8	4.8	4.8	4.8	4.7	4.7	4.6	4.6	4.6
400	*****	4.1	4.1	4.1	4.1	4.0	4.0	4.0	4.0	4.0	4.0	4.0
500	*****	3.8	3.8	3.8	3.8	3.8	3.7	3.6	3.5	3.4	3.3	3.3
750	*****	*****	*****	2.9	2.8	2.7	2.7	2.7	2.7	2.7	2.6	2.6
1000	*****	*****	*****	2.3	2.2	2.2	2.2	2.2	2.2	2.2	2.2	2.2
1500	*****	*****	*****	*****	*****	1.6	1.6	1.6	1.6	1.6	1.6	1.6
2000	*****	*****	*****	*****	*****	*****	1.2	1.2	1.2	1.2	1.2	1.2
2500	*****	*****	*****	*****	*****	*****	*****	0.9	0.9	0.8	0.8	0.8
3000	*****	*****	*****	*****	*****	*****	*****	*****	0.6	0.5	0.4	0.4

11 Questionnaire, dictionnaire de données et feuilles de codage

Veillez consulter les fichiers ci-dessous aux fins de l'Enquête auprès des titulaires d'un doctorat 2003-2004.

Questionnaires:

Enquête titulaire dun doctorat-F.pdf
Survey of Earned Doctorates-E.pdf

Dictionnaire de données et feuilles de codage

Classification des programmes d'enseignement (CPE)

Annexe A - CPE.pdf

Système de classification de l'industrie de l'Amérique du Nord (SCIAN) 2002

Annexe B – SCIAN.pdf OU Annexe B – SCIAN.doc

Classification nationale des professions pour statistiques (CNP-S) de 2001

Annexe C – SCP-S.pdf OU Annexe C – SCP-S.doc

12. Cliché d'enregistrement à valeurs univariées

Voir ETD2003-2004LvCds_F.pdf OU ETD2003-2004LvCds_F.doc pour le cliché d'enregistrement à chiffres univariés.

