

ENQUÊTE SUR LA PROMOTION DE LA SANTÉ, 1990

Guide de l'utilisateur des microdonnées

Table des matières

| | | |
|------------|-----------------------------------------------------------------|-----------|
| 1.0 | Introduction | 1 |
| 2.0 | Contexte | 3 |
| 3.0 | Objectifs | 5 |
| 4.0 | Concepts/Définitions | 7 |
| 5.0 | Plan de l'enquête | 9 |
| 5.1 | Champ de l'enquête | 9 |
| 5.2 | Stratification | 9 |
| 5.3 | Plan de sondage | 10 |
| 5.4 | Taille de l'échantillon par province | 12 |
| 6.0 | Méthodologie de la collecte des données | 13 |
| 6.1 | Conception du questionnaire | 13 |
| 6.2 | Méthodologie de la collecte | 14 |
| 6.3 | Période de collecte | 14 |
| 6.4 | Interview | 14 |
| 7.0 | Traitement des données | 15 |
| 7.1 | Saisie des données | 15 |
| 7.2 | Vérification | 15 |
| 7.3 | Variables dérivées | 16 |
| 7.4 | Pondération | 16 |
| 8.0 | Erreur d'échantillonnage | 17 |
| 8.1 | Coefficients de variation pour les estimations par catégorie | 18 |
| 8.2 | Exemples | 21 |
| 8.3 | Coefficients de variation d'estimations quantitatives | 22 |
| 8.4 | Intervalles de confiance | 22 |
| 9.0 | Erreur non due à l'échantillonnage | 25 |
| 9.1 | Non-réponse complète | 25 |
| 9.2 | Non-réponse partielle | 26 |
| 9.3 | Taux de réponse | 26 |
| 9.4 | Couverture | 26 |
| 9.5 | Sous-estimation | 27 |

| | | |
|-------------|--------------------------------------------------------------------------|-----------|
| 10.0 | Lignes directrices relatives à la publication et à la diffusion | 29 |
| 10.1 | Règles relatives à l'arrondissement | 29 |
| 10.2 | Règles relatives à la pondération de l'échantillon dans les applications | 30 |
| 10.3 | Règles relatives à la variabilité d'échantillonnage | 31 |
| 10.4 | Règles relatives à l'analyse statistique | 32 |
| 11.0 | Questionnaires | 33 |
| 12.0 | Méthodes de pondération | 35 |
| 12.1 | Estimation | 35 |
| 12.2 | Pondération de l'Enquête Promotion Santé | 35 |
| 12.3 | Politique concernant la pondération | 39 |
| 13.0 | Cliché d'articles | 41 |
| 14.0 | Notes explicatives sur le cliché d'articles | 43 |
| 14.1 | Sigle | 43 |
| 14.2 | Codification | 43 |
| 14.3 | Codes (valeurs) | 44 |
| 15.0 | Tableaux de la variabilité d'échantillonnage approximative | 45 |

1.0 Introduction

L'enquête sur la promotion de la santé (EPS) a été menée au mois de juin 1990. Elle a été réalisée par Statistique Canada, pour le compte de Santé et Bien-être social Canada.

Le présent guide a été élaboré pour faciliter la manipulation du fichier de microdonnées contenant les résultats de l'enquête. Toutes les questions relatives au fichier ou à son utilisation doivent être adressées au :

Services aux clients
Division des enquêtes spéciales
Statistique Canada
Tél: (613) 951-7355 OU 1-888-297-7355
Télécopieur: (613) 951-3012
Courrier: des@statcan.ca

IL EST IMPORTANT POUR LES UTILISATEURS DE SE FAMILIARISER AVEC LE CONTENU DU PRÉSENT DOCUMENT AVANT DE PUBLIER OU DE DIFFUSER DES ESTIMATIONS PRODUITES À PARTIR DU FICHER DE MICRODONNÉES DE L'ENQUÊTE SUR LA PROMOTION DE LA SANTÉ AU CANADA.



2.0 Contexte

En 1981, le gouvernement du Canada a donné à la Direction de la promotion de Santé et Bien-être social Canada le mandat de mettre en oeuvre un programme national de promotion de la santé. Une des composantes de ce programme était d'élaborer une enquête nationale sur l'existence et la répartition d'une variété d'habitudes de santé qui influent sur l'avenir des canadiens aux niveaux de la promotion de la santé et de la prévention des maladies.

La planification pour l'«Enquête promotion santé Canada» a débuté en 1982 et s'est soldée par une enquête nationale menée en juin 1985 par Statistique Canada et touchant 11 181 canadiens adultes. L'objectif de l'EPS de 1985 était de créer une base de données nationale sur les habitudes et les conditions de santé qui sont étroitement liées à la santé et au bien-être de la population adulte canadienne. L'EPS a permis d'obtenir des données de base tant nationales que provinciales sur les attitudes, les croyances, les intentions et les comportements des canadiens adultes en ce qui a trait à une variété de questions liées à la promotion de la santé (condition physique, nutrition, sécurité et usage du tabac, alcool et drogues).

Afin de mettre à jour la base de données de l'EPS, un second cycle de l'enquête a été mené en juin 1990. Comme en 1985, Santé et Bien-être social Canada parraine l'enquête et Statistique Canada assure la collecte des données.

L'EPS a été menée en vertu de la Loi sur la statistique, Lois révisées du Canada, 1985, Chapitre S19. La collecte des données est enregistrée au numéro d'inscription SQC/ENM-040-03909. C'est une enquête à participation volontaire, et elle a reçu l'approbation du ministre exigée aux termes de l'article 8 de la Loi sur la statistique.



3.0 Objectifs

Les objectifs de l'enquête Promotion de la santé (1990) étaient de mettre à jour et de développer des données de base nationales et provinciales sur les connaissances, les attitudes, les croyances, les intentions et les comportements des Canadiens adultes en ce qui concerne une variété de questions liées à la promotion de la santé.

Parmi les sujets traités, on retrouve la perception de la santé, la tension artérielle et le cholestérol, la consommation d'alcool, l'usage du tabac, l'exercice, l'alimentation, la dentition, le milieu de travail et la sécurité, l'environnement et les maladies transmises sexuellement.

Dans le cadre du cycle de 1990, il sera possible de comparer les résultats d'enquêtes précédentes (c.-à-d. l'Enquête santé Canada de 1978 et l'EPS de 1985) par rapport à nombre de questions liées à la promotion de la santé. On figurera ensuite graphiquement les tendances nationales au cours des années afin d'évaluer leurs effets sur les programmes de promotion de la santé et de prévention des maladies de la Direction de la promotion de la santé.



4.0

Concepts/Définitions

L'EPS est une enquête téléphonique, on a donc essayé d'utiliser une terminologie facile à comprendre et ainsi minimiser de très longues explications données au téléphone.

Cependant, quelques définitions de base ont du être données aux répondants afin de normaliser les réponses. Les définitions suivantes sont indiquées sur le questionnaire des données de l'EPS et ont été lues à tous les répondants.

Taux de cholestérol sanguin élevé :

Si un professionnel de la santé a déjà indiqué au répondant que son taux de cholestérol sanguin est élevé, alors cela voulait dire que le répondant a eu une prise de sang pour mesurer son taux de cholestérol sanguin.

Exercice :

Par exercice, on voulait dire des activités vigoureuses telles que l'aérobie, le jogging, les sports d'équipe, les cours de danse ou la marche vive.

Un verre d'alcool :

Un verre signifiait :

- une bouteille de bière ou un verre de bière en fût;
- un petit verre de vin; ou
- une once et demie de boisson forte, diluée ou non.

Relations sexuelles :

Les relations sexuelles incluait les relations homosexuelles et hétérosexuelles.

Canadien Français :

Si la réponse à la question R5 était Canadien Français, Québécois, Acadien ou Franco-Ontarien, alors les deux catégories Français et Canadien étaient cochées.



5.0

Plan de l'enquête

Comme l'enquête de 1985, le plan de sondage de ce deuxième cycle est créé par le système de composition aléatoire (SCA), technique par laquelle les numéros de téléphone sont générés au hasard par ordinateur et sont ensuite composés systématiquement.

Se fondant sur l'expérience de l'EPS de 1985, Statistique Canada n'a pas effectué de sondage téléphonique au hasard dans le Nord lors du cycle de 1990. Une collecte distincte des données, à partir de méthodes reflétant mieux la réalité du Nord du Canada, est considérée pour le Yukon et les Territoires du Nord-Ouest. Santé et Bien-être social Canada travaille de concert avec les administrations territoriales, la question de logistique de cette enquête.

5.1

Champ de l'enquête

Toute personne âgée de 15 ans et plus, vivant au Canada, faisait parti de la population cible de l'EPS, exception faite des deux cas suivants :

1. résidents du Yukon et des Territoires du Nord Ouest;
2. résidents à plein temps d'institutions.

Par sa technique d'échantillonnage, l'EPS est une enquête téléphonique. Par conséquent, les ménages et, les personnes vivant dans ce ménage, n'ayant pas le téléphone ont été exclus de l'enquête. Ceci représente moins de 3 % de la population totale. Cependant, lors de la pondération des données, on a tenu compte des personnes n'ayant pas le téléphone.

5.2

Stratification

En prévision du prélèvement de l'échantillon, chacune des dix provinces a été divisée en strates ou régions géographiques. De façon générale, pour chaque province une strate représentait les régions métropolitaines de recensement (RMR) de la province et l'autre, les régions qui ne sont pas des RMR. Puisque l'Ontario et la Saskatchewan sont tous deux échantillonnées à partir de deux bureaux régionaux, un plus grand nombre de strates étaient intégrées au plan de sondage pour ces régions.

5.3 Plan de sondage

L'échantillon de l'enquête est créé par le système de composition aléatoire (SCA) qui utilise deux méthodes pour générer des numéros de téléphone : la méthode Waksberg et la méthode des banques non valides (RBNV).

5.3.1 La méthode Waksberg :

La méthode Waksberg a été utilisée dans deux provinces: l'Île-du-Prince-Édouard et la strate hors des RMR du Québec. La méthode est basée sur un plan de sondage à deux degrés qui augmente la probabilité de rejoindre des ménages. Les lignes qui suivent, décrivent l'application de la méthode à ces deux strates pour l'EPS de 1990.

Une liste à jour de toutes les combinaisons de codes régionaux et d'indicatifs a été obtenue. On a ensuite ajouté à ces séries toutes les combinaisons possibles de deux chiffres (c-à-d qu'on a repéré toutes les banques de numéros de 100 numéros consécutifs qui peuvent exister pour chaque combinaison de code régional et d'indicatif). On a ainsi obtenu, pour chaque strate, une liste de toutes les combinaisons possibles des huit premiers chiffres d'un numéro de téléphone qui en compte dix. Ces numéros de huit chiffres formaient les unités de sondage du premier degré (ou les unités primaires d'échantillonnage – UPÉ).

Dans chaque strate, un de ces numéros de huit chiffres a été choisi au hasard et les deux derniers chiffres ont ensuite été générés de façon aléatoire. On a ensuite composé le numéro ainsi obtenu (appelé numéro primaire) pour vérifier s'il correspondait bien à celui d'un ménage (c-à-d un numéro qui ne correspondait pas à une entreprise commerciale, une institution, etc.):

- S'il ne permettait pas de rejoindre un ménage, il était rejeté immédiatement.
- Dans le cas contraire, d'autres numéros, appelés numéros secondaires, étaient générés dans la même banque. On a également composé ces numéros pour vérifier s'ils permettaient de rejoindre un ménage.

Des numéros secondaires ont été générés sans interruption jusqu'à ce que (1) cinq ménages additionnels puissent être rejoints dans chaque banque choisie ou que (2) la banque soit épuisée ou encore que (3) la période d'enquête soit terminée.

Les numéros primaires et secondaires ont été continuellement générés durant la période d'enquête jusqu'à ce qu'un nombre prédéterminé de ménages ait été atteint pour chacune des deux strates où la méthode Waksberg a été utilisée.

Dans chaque ménage rejoint, on a essayé de mener une entrevue auprès d'une personne choisie de façon aléatoire parmi les membres du ménage.

Le principe sous-jacent à cette technique est que si l'on produit des numéros de téléphone au hasard, on réussit mieux et plus efficacement à rejoindre des numéros de téléphone résidentiels si les numéros ainsi produits sont groupés dans des banques pour lesquelles on a constaté qu'il y a au moins un numéro résidentiel. Pour l'EPS 1990, 53.7 % des numéros secondaires composés correspondaient à une résidence comparé à seulement 14.6 % des numéros primaires.

5.3.2 Méthode d'Élimination des banques non valides (EBNV)

La méthode de l'élimination des banques non valides est une méthode de composition de numéros de téléphone au hasard dans laquelle on tente de repérer toutes les banques valides dans une région donnée, c-à-d de repérer toutes les banques contenant au moins un ménage. Par conséquent, tous les numéros de téléphone dans les banques non valides sont supprimées de la base de sondage. Cette méthode a été utilisée pour produire l'échantillon dans toutes les strates sauf celles de l'Île-du-Prince-Édouard et de la région hors des RMR de Québec. Les lignes qui suivent, décrivent l'application de la méthode à ces strates pour l'EPS de 1990.

Une liste de toutes les banques qui contenaient au moins un ménage a été fournie par les compagnies de téléphone à travers le Canada. Après avoir attribuée chaque banque à une strate, on a généré systématiquement un échantillon de numéros de téléphone lors de la première journée d'entrevues.

Les numéros de téléphone ont été composés pour vérifier s'il correspondait bien à celui d'un ménage. Comme avec la méthode Waksberg, on a essayé de mener une entrevue auprès d'une personne choisie de façon aléatoire parmi les membres du ménage.

Pour l'EPS 1990, avec la méthode EBNV, 52.9 % des numéros composés correspondaient à une résidence.

5.4

Taille de l'échantillon par province

| PROVINCE | ÉCHANTILLON ALLOUÉ | ENREGISTREMENTS SUR LE FICHER |
|-----------------------|--------------------|-------------------------------|
| Terre-Neuve | 1 000 | 1 088 |
| Île-du-Prince-Édouard | 1 000 | 942 |
| Nouvelle-Écosse | 1 000 | 1 002 |
| Nouveau-Brunswick | 1 000 | 943 |
| Québec | 1 880 | 1 728 |
| Ontario | 2 230 | 2 280 |
| Manitoba | 1 000 | 1 066 |
| Saskatchewan | 1 000 | 990 |
| Alberta | 2 581* | 2 530* |
| Colombie-Britannique | 1 250 | 1 223 |
| CANADA | 13 941 | 13 792 |

* Échantillon élargi par l'achat d'échantillons supplémentaires par l'alberta.

6.0

Méthodologie de la collecte des données

6.1

Conception du questionnaire

Les deux principaux outils pour mener l'enquête étaient la formule de contrôle et le questionnaire de l'EPS 1990 (reférez-vous à la section 11 pour le détail des questions).

La formule de contrôle est utilisée pour la sélection du répondant à l'intérieur du ménage. Les questions sont choisies conformément aux enquêtes menées selon la méthode de composition aléatoire par téléphone.

Bien que le questionnaire de 1990 ressemble étroitement à celui de 1985, plusieurs modifications y ont été apportées :

- 1) Plusieurs questions ont été reformulées pour corriger des lacunes observées. La question I1 (portant sur le nombre de petits déjeuners) en est un exemple.
- 2) Certaines questions demandées en 1985 ne se retrouvent pas dans l'enquête de 1990. Le but étant de diminuer un dédoublement de questions avec d'autres enquêtes sur la santé et pour tenir davantage compte du nouvel énoncé de mission de la Direction de la promotion de la santé.
- 3) De nouvelles sections ont été ajoutées, entre autres la prévention des maladies transmises sexuellement et l'hygiène dentaire.

Avant de mener l'enquête nationale, un test des deux questionnaires de l'EPS 1990 a été mené à Halifax et à Montréal, deux bureaux régionaux de Statistique Canada, en mars 1990. Près de 300 répondants ont été interviewés dans chacun des deux bureaux. Le test consistait à vérifier la qualité des instruments de collecte, dans les deux langues officielles (c-à-d la durée de l'entrevue, la réaction des répondants, etc.), auprès des répondants répartis dans quatre strates, soit deux strates, une rurale et une urbaine, dans chacune des deux provinces (Nouvelle-Écosse et Québec). Des modifications ont été apportées à la formulation des questions suite à l'analyse des résultats du test ainsi qu'aux commentaires des intervieweurs.

6.2

Méthodologie de la collecte

La formule de contrôle est utilisée chaque fois qu'un numéro de téléphone est composé par un intervieweur. Le but du formulaire est, tout d'abord, de déterminer si le numéro de téléphone composé appartient ou non à un ménage. Dans l'affirmative, on liste les membres du ménage sur le formulaire. Ensuite, on échantillonne au hasard, à l'aide d'une grille de sélection pré-imprimée, un membre du ménage âgé de 15 ans et plus.

Par la suite, on continue l'entrevue téléphonique avec la personne choisie. Si cette personne n'était pas disponible sur le moment pour répondre au questionnaire sur la Promotion de la santé-1990, un rendez-vous était fixé pour la rappeler plus tard. Toutes les entrevues ont été menées avec la personne sélectionnée, sans approximation; la raison étant la subjectivité et l'aspect personnel des questions de l'enquête.

6.3

Période de collecte

Les entrevues ont été menées à partir des huit bureaux régionaux de Statistique Canada durant tout le mois de juin 1991.

Tous les entrevues ont été menées entre 8.30 et 21.30, heure local, du lundi au vendredi. Des entrevues ont aussi été menées le samedi durant la journée.

6.4

Interview

La collecte de données a été menée par des intervieweurs, hautement qualifiés, de Statistique Canada. Toutes les questions ont été posées au répondant conformément à des instructions strictes. Leur formation a été particulièrement axée sur la façon de rester neutre et de s'en tenir strictement aux directives. On leur a souligné la nature délicate de certaines sections du questionnaire et ils ont été formés pour en tenir compte. Le manuel de formation et le manuel de l'intervieweur ont été fournis aux intervieweurs. L'intervieweur senior a aussi reçu le manuel des procédures.

Les interviews ont été menés dans une des deux langues officielles. Si un répondant ne pouvait communiquer dans aucune des deux langues, l'entrevue téléphonique n'avait pas lieu.

7.0 Traitement des données

7.1 Saisie des données

La saisie des données de l'EPS a été effectuée, aux mois de juin et juillet 1990, à l'aide de mini-ordinateurs dans chacun des huit bureaux régionaux de Statistique Canada. Le programme de saisie des données acceptait un certain nombre de codes pour chaque question et il respectait automatiquement l'enchaînement voulu des questions. Les données ont été transmises à Ottawa par voie électronique en vue de la création d'un premier fichier informatique.

7.2 Vérification

Lors de la première étape du traitement, on a créé un fichier de 13 960 enregistrements. Pour fin de compatibilité avec la plupart des logiciels statistiques, une valeur numérique a été associée à tous les champs du fichier.

Un logiciel sur mesure a été développé pour corriger les incohérences des données dues à des erreurs d'entrevues ou de saisies. La vérification des données a été ascendante, c-à-d qu'à l'intérieur d'un intervalle logique, on a repéré les réponses afin de corriger l'enchaînement voulu des questions. Par la suite, une valeur numérique, uniforme à l'intérieur du fichier (8, 98, etc.), a été assigné aux questions, qui ne demandait pas de réponses.

De plus, on a développé des contrôles entre divers questions. Par exemple, on a vérifié l'âge avec la date de naissance pour s'assurer de l'uniformité des réponses.

Les codes d'industrie et d'occupation ont été assignés par le logiciel de codification automatique de l'enquête sur la population active.

À la fin du processus de vérification, le fichier microdonnées contenait 13 792 enregistrements. (N.B.: 268 enregistrements ont été retranchés du fichier pour différentes raisons. Reférez-vous à la section 9.2 et 9.3)

7.3 Variables dérivées

Suite à tous les contrôles numériques, on a dérivé des variables pour répondre aux besoins des utilisateurs. Comme exemple, on peut nommer la taille du ménage (DVHSIZE), l'indice du poids corporel (BMI) et la province (DVPROV).

Pour fin de comparaison avec d'autres fichiers et d'après les exigences du comité des fichiers microdonnées, deux groupes de codes ont été dérivés pour chacun des codes d'activités économiques et d'occupation.

7.4 Pondération

L'estimation d'un échantillon aléatoire comme celui de l'EPS repose sur un principe selon lequel chaque personne de l'échantillon représente, en plus d'elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple de 2 % de la population, chaque personne de l'échantillon représente 50 personnes de la population.

L'étape de la pondération est une étape où l'on calcule, pour chaque enregistrement du fichier de microdonnées, ce qu'est ce nombre, c'est-à-dire le poids de l'enregistrement, que l'on inscrit dans le fichier. On doit utiliser ce poids pour produire des estimations à partir du fichier de microdonnées. Par exemple, si l'on veut estimer le nombre de personnes qui ont partiellement terminées leurs études primaires, on sort les enregistrements des personnes qui présentent cette caractéristique et on additionne leurs poids.

On trouvera au chapitre 12 des précisions sur la méthode utilisée pour calculer les poids d'échantillonnage.

8.0

Erreur d'échantillonnage

Les estimations produites à partir des résultats de cette enquête sont basées sur des données recueillies auprès d'un échantillon de personnes. On aurait peut-être obtenu des estimations quelque peu différentes si on avait effectué un recensement en utilisant, notamment, le même questionnaire, les mêmes intervieweurs et surveillants et les mêmes méthodes de traitement. La différence entre les estimations basées sur un échantillon et celles résultant d'un dénombrement complet entrepris dans des conditions semblables s'appelle l'erreur d'échantillonnage des estimations.

Comme il est inévitable que les estimations établies à partir d'un sondage soient entachées d'une erreur d'échantillonnage, il convient que les chercheurs fournissent aux utilisateurs une indication de l'ampleur de cette erreur. Dans cette section, nous donnons un aperçu des mesures de l'erreur d'échantillonnage dont on se sert couramment à Statistique Canada et dont on invite les utilisateurs à se servir aussi s'ils comptent produire des estimations à partir de ce fichier de microdonnées.

Pour mesurer l'ampleur éventuelle de l'erreur d'échantillonnage, on se base sur l'écart-type (et) des estimations produites à partir des résultats de l'enquête. Cependant, comme il est possible de produire une grande variété d'estimations à partir des résultats d'une enquête comme celle-ci, on exprime habituellement l'écart-type d'une estimation par rapport à l'estimation en question. Cette mesure, appelée coefficient de variation (cv) de l'estimation, s'obtient en divisant l'écart-type de l'estimation par l'estimation elle-même et s'exprime en pourcentage de l'estimation.

$$cv_x = et_x / \bar{X}$$

Par exemple, supposons que, d'après les résultats de l'enquête, on estime que 25 % de tous les adultes (personnes âgées de 15 et plus) disent qu'ils ont une excellente santé et que l'écart-type de l'estimation soit de 0.012. Le cv de l'estimation se calcule alors ainsi :

$$cv_x = et_x / \bar{X} = 0.012 / 0.25 = 0.048 = 4.8 \%$$

N.B.: Veuillez consulter le chapitre 10.3 au sujet des règles relatives à la variabilité de l'échantillonnage.

Avant de parler de la façon d'obtenir ces mesures, il est utile de décrire les deux principaux types d'estimations ponctuelles des caractéristiques démographiques qu'il est possible de produire à partir des fichiers de microdonnées de l'EPS.

(1) Estimations par catégorie

Les estimations par catégorie sont des estimations de l'effectif, de la proportion ou du pourcentage de la population interrogée présentant certaines caractéristiques ou se situant dans une catégorie donnée. Comme exemples de ces estimations, mentionnons le nombre de personnes de 15 à 24 ans qui sont en «excellente santé» ou la proportion de la population de Nouvelle-Écosse constituée de femmes en «mauvaise santé».

Dans ce contexte, une estimation du nombre de personnes présentant une certaine caractéristique est appelée estimation d'un agrégat.

(2) Estimations quantitatives

Les estimations quantitatives sont des estimations des totaux ou des moyennes, et autres mesures de tendance centrale basées sur les réponses de certains ou de la totalité des membres de la population interrogée. Ces estimations sont plus précisément des estimations de forme X/\hat{Y} , X étant une estimation de l'effectif total de la population interrogée et \hat{Y} , une estimation du nombre de personnes dans la population interrogée contribuant à ce total.

Dans la présente enquête, le nombre moyen de partenaires, lors de relations sexuelles, au cours des douze derniers mois pour des Canadiens âgés de 15 ans et plus est une estimation quantitative.

8.1

Coefficients de variation pour les estimations par catégorie

Comme nous voulions fournir aux utilisateurs des cv qui soient applicables à un large éventail d'estimations par catégorie établies à partir de ce fichier de microdonnées et auxquelles ils puissent avoir accès facilement, nous avons produit une série de tableaux de référence, appelés Tableaux de la variabilité d'échantillonnage approximative, qui se trouvent au chapitre 15.

On a calculé ces cv à l'aide de la formule de la variance s'appliquant à l'échantillonnage aléatoire simple, à laquelle on a ajouté un facteur qui rend compte du fait qu'il s'agit d'échantillonnage en grappes à plusieurs degrés. Ce facteur, qu'on appelle l'effet du plan, a été déterminé en calculant d'abord les effets dus au plan pour une vaste gamme de caractéristiques et en choisissant ensuite parmi les valeurs obtenues une valeur raisonnable qui ne donnera pas une fausse impression de grande précision. On peut obtenir de Statistique Canada, moyennant remboursement des coûts, des estimations de la variance réelle associée à des variables précises.

Les règles suivantes devraient permettre à l'utilisateur de trouver, dans les tableaux de la variabilité d'échantillonnage approximative, les cv correspondant aux estimations de l'effectif, de la proportion ou du pourcentage de la population interrogée présentant une certaine caractéristique, ainsi que des rapports et des différences entre les estimations.

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

Le cv dépend uniquement de l'ordre de grandeur de l'estimation elle-même. Dans le tableau de la variabilité d'échantillonnage approximative correspondant à la région géographique qui vous intéresse, trouvez le nombre estimé dans la colonne le plus à gauche du tableau (intitulée "Numérateur du pourcentage") et suivez les astérisques (s'il y en a) jusqu'au premier chiffre rencontré. Ce chiffre est le cv approximatif.

Règle 2 : Estimations des proportions ou pourcentages de personnes présentant une caractéristique

Le cv d'une proportion ou pourcentage estimé dépend à la fois de l'ordre de grandeur de la proportion ou du pourcentage et de l'ordre de grandeur du total sur lequel est basé ce pourcentage. Les proportions ou pourcentages estimés sont relativement plus fiables que les estimations correspondantes du numérateur de la proportion ou du pourcentage, particulièrement si la proportion ou le pourcentage est basé sur un sous-ensemble de la population globale. (Notez que, dans les tableaux, les cv diminuent de valeur de gauche à droite.)

Lorsque la proportion ou le pourcentage est basé sur la population totale de la région géographique à laquelle se rapporte le tableau, le cv de la proportion ou du pourcentage est le même que celui du numérateur de la proportion ou du pourcentage. Dans ce cas, on peut appliquer la règle 1.

Lorsque la proportion ou le pourcentage est basé sur un sous-ensemble de la population globale (par exemple, les personnes d'un groupe d'âge particulier), on doit se reporter à la proportion ou au pourcentage (de gauche à droite en haut du tableau) et au numérateur de la proportion ou du pourcentage (de haut en bas dans la partie gauche du tableau). Le cv se trouve à l'intersection de la ligne et de la colonne appropriées.

Règle 3 : Estimations de l'écart entre les agrégats et les pourcentages

On indique le cv de la différence entre deux estimations par l'écart-type de la différence divisé par la différence. L'écart-type de la différence entre deux estimations est approximativement égal à la racine carrée de la somme des carrés de chaque écart-type considéré séparément.

Avec $\bar{d} = \bar{X}_1 - \bar{X}_2$

$$et_{\bar{d}} = \sqrt{(\bar{X}_1 * \alpha_1)^2 + (\bar{X}_2 * \alpha_2)^2}$$

$$cv_{\bar{d}} = et_{\bar{d}} / \bar{d}$$

Où: \bar{X}_1 et \bar{X}_2 sont les estimations.
 α_1 et α_2 sont les cv de \bar{X}_1 et \bar{X}_2 respectivement.

Cette formule est précise pour la différence entre des estimations distinctes et non corrélées; autrement, elle n'est qu'approximative.

Règle 4 : Estimations des rapports

Lorsque le numérateur est un sous-ensemble du dénominateur, le rapport doit être exprimé en pourcentage et la règle 2 s'applique. C'est ce qu'on ferait, par exemple, si le dénominateur est le nombre d'hommes et le numérateur, le nombre d'hommes ayant une caractéristique unique.

Lorsque le numérateur n'est pas un sous-ensemble du dénominateur, le cv du rapport de deux estimations ($R = \bar{X}_1/\bar{X}_2$) est égal à l'écart-type du rapport divisé par le rapport ($cv_R = st_R / R$). L'écart-type du rapport des deux estimations est approximativement égal à la racine carrée de la somme des carrés de chaque cv, considéré séparément, multiplié par le rapport. Autrement dit, le cv du rapport est approximativement égal à la racine carrée de la somme des carrés de chaque cv considéré séparément:

Avec $R = \bar{X}_1 / \bar{X}_2$

$$et_R = R * \sqrt{\alpha_1^2 + \alpha_2^2}$$

$$cv_R = et_R / R$$

$$cv_R = (R * \sqrt{\alpha_1^2 + \alpha_2^2}) / R$$

$$cv_R = \sqrt{\alpha_1^2 + \alpha_2^2}$$

où: \bar{X}_1 et \bar{X}_2 sont les estimations.
 α_1 et α_2 sont les cv de \bar{X}_1 et \bar{X}_2 respectivement.

La formule aura tendance à surestimer l'erreur si \bar{X}_1 et \bar{X}_2 sont en corrélation positive et à la sous-estimer si \bar{X}_1 et \bar{X}_2 sont en corrélation négative.

Règle 5 : Estimations de la différence entre deux rapports

Dans ce cas, les règles 3 et 4 sont combinées. On détermine d'abord les cv pour les deux rapports à l'aide de la règle 4, et ensuite le cv de la différence entre les deux à l'aide de la règle 3.

8.2 Exemples

Voici deux exemples basés sur le fichier microdonnées de l'EPS qui aideront l'utilisateur à appliquer les règles précédentes. Reférez-vous au chapitre 13 pour les noms des variables et leur codes.

Exemple 1

Supposons que l'utilisateur estime, d'après le fichier de microdonnées, que 200 260 adultes ne mangent rien au petit déjeuner (c-à-d qu'ils ont pris du café ou du thé ou qu'ils n'ont rien pris du tout durant la période de référence; I1A= 7). Comment déterminer le cv de cette estimation?

- 1) On se reporte au tableau pour le Manitoba.
- 2) L'agrégat estimé, 200 260, ne figure pas dans la colonne de gauche (celle du "Numérateur du pourcentage"), de sorte qu'il est nécessaire de prendre le chiffre le plus proche, soit 200 000.
- 3) Pour trouver le cv d'un agrégat estimé, il faut se reporter à la première inscription sans astérisque de cette ligne, soit 6.1 %.
- 4) Le cv approximatif de l'estimation est donc 6.1 %.

Exemple 2

Supposons que l'utilisateur estime alors que 40.6 % (ou 81 208) du 200 260 adultes, qui sautent le petit déjeuner au Manitoba, ont un indice de masse corporel (BMI) entre 20.0 et 24.9. Comment déterminer le cv de cette estimation?

- 1) On se reporte au tableau pour le Manitoba.
- 2) L'estimation étant un pourcentage basé sur un sous-ensemble de la population totale (c-à-d les personnes ayant sauté le petit déjeuner), il est nécessaire d'utiliser à la fois le pourcentage (33 %) et le numérateur du pourcentage (81 208) pour déterminer le cv.
- 3) Le numérateur, 81 208, ne figure pas dans la colonne de gauche (celle du "Numérateur du pourcentage"), de sorte qu'il est nécessaire d'utiliser le chiffre le plus proche, soit 80 000. L'estimation du pourcentage n'apparaît elle non plus dans aucune des têtes de colonne, de sorte qu'il faut encore utiliser le chiffre le plus proche, soit 40.0 %.

- 4) Le chiffre se trouvant à l'intersection de la ligne et de la colonne utilisées, soit 8.6 % est le cv à retenir.
- 5) Le cv approximatif du pourcentage estimé est donc 8.6 %.

8.3

Coefficients de variation d'estimations quantitatives

Les variables importantes qui nous intéressent dans l'EPS se rapportent à des catégories. Pour ce qui est des estimations quantitatives, il faudra construire des tableaux spéciaux sur demande pour déterminer l'erreur d'échantillonnage dont elles sont entachées.

Cependant, comme règle, si le nombre pondéré de personnes sur lequel le total quantitatif estimé est fondé n'est pas publiable, alors ce total n'est pas publiable non plus. Cela signifie que pour toute tabulation comportant des estimations quantitatives, on devrait aussi produire des tableaux donnant le nombre estimatif de personne dans chaque case.

8.4

Intervalles de confiance

Bien que les cv soient couramment utilisés, l'intervalle de confiance d'une estimation est une mesure de l'erreur d'échantillonnage qui se comprend mieux intuitivement.

L'intervalle de confiance indique la probabilité que la valeur réelle de la population se situe entre deux valeurs donnés. Par exemple, on peut décrire comme suit un intervalle de confiance à 95 %.

Si on prélève un nombre indéfini d'échantillons de la population et qu'à chaque échantillon l'intervalle de confiance d'une estimation est d'une étendue différente, dans 95 % des échantillons la valeur réelle de la population se situera dans l'intervalle de confiance.

En utilisant l'écart-type, on peut obtenir les intervalles de confiance d'une estimation en vertu de l'hypothèse que, à chaque échantillonnage de la population, les diverses estimations obtenues pour une caractéristique démographique sont distribuées normalement autour de la valeur réelle. Selon cette hypothèse, il y a une probabilité d'environ 68 % que la différence entre une estimation d'échantillon et la valeur réelle de la population soit inférieure à un écart-type, d'environ 95 % que la différence soit inférieure à deux écarts-types, et d'environ 99 % que la différence soit inférieure à trois écarts-types. Ces différents degrés de confiance sont appelés «niveaux de confiance».

Pour une estimation \bar{X} , on exprime généralement les intervalles de confiance au moyen de deux nombres, l'un au-dessous et l'autre au-dessus de l'estimation, sous la forme $(\bar{X}-k, \bar{X}+k)$ où k est déterminé selon le niveau de confiance désiré et l'erreur d'échantillonnage de l'estimation.

On peut calculer directement les intervalles de confiance d'une estimation à partir des Tableaux de la variabilité d'échantillonnage en trouvant d'abord dans le tableau approprié le cv de l'estimation \bar{X} , et en utilisant ensuite la formule suivante pour convertir ce dernier en intervalle de confiance :

$$IC_{\bar{X}} = \{\bar{X} - (t * \bar{X} * \alpha), \bar{X} + (t * \bar{X} * \alpha)\}$$

où: α est le coefficient de variation de \bar{X}

$t = 1.0$ si on désire un intervalle de confiance à 68 %

$t = 1.6$ si on désire un intervalle de confiance à 90 %

$t = 2.0$ si on désire un intervalle de confiance à 95 %

$t = 3.0$ si on désire un intervalle de confiance à 99 %

Exemple

On calculerait de la manière suivante l'intervalle de confiance à 95 % de la proportion estimée d'adultes ayant sautés le petit déjeuner au Manitoba et ayant un indice de masse corporel entre 20.0 et 24.9 (d'après l'exemple 2 de la section 8.2) :

$$\bar{X} = 40.6 \% \text{ (ou, sous forme de proportion} = 0.406)$$

$$t = 2$$

$$cv_{\bar{X}} = 8.6 \% \text{ (0.086 exprimé en proportion)}$$

$$IC_{\bar{X}} = \{0.406 - (2 * 0.406 * 0.086), 0.406 + (2 * 0.406 * 0.086)\}$$

$$IC_{\bar{X}} = \{0.406 - 0.0698, 0.406 + 0.0698\}$$

$$IC_{\bar{X}} = \{0.3362, 0.4758\}$$

On peut dire avec 95 % de chances de ne pas se tromper qu'entre 33.6 % et 47.6 % des personnes qui sautent le petit déjeuner ont un indice de masse corporel entre 20.0 et 24.9.



9.0

Erreur non due à l'échantillonnage

Des erreurs non dues à l'échantillonnage peuvent se produire à presque toutes les étapes d'une enquête. Les intervieweurs peuvent mal interpréter les directives, les répondants peuvent se tromper en répondant aux questions, les réponses peuvent être inscrites incorrectement sur le questionnaire et des erreurs peuvent être commises au moment du traitement et de la totalisation des données. Ce sont là des exemples d'erreurs non dues à l'échantillonnage.

Si le nombre d'observations est élevé, les erreurs qui surviennent au hasard auront généralement peu d'effet sur les estimations produites à partir des résultats d'une enquête. En revanche, les erreurs qui se produisent systématiquement contribueront à biaiser les estimations. Dans l'EPS, on n'a ménagé ni le temps ni les efforts pour réduire au maximum les erreurs non dues à l'échantillonnage. On a appliqué des mesures d'assurance de la qualité à chaque étape de la collecte et du traitement des données afin de contrôler la qualité de celles-ci. Parmi ces mesures, mentionnons l'utilisation d'intervieweurs hautement qualifiés, la prestation aux intervieweurs d'une formation poussée sur les procédures et le questionnaire de l'EPS, l'observation des intervieweurs pour voir si le questionnaire avait été mal conçu ou les directives mal interprétées, l'application de procédures visant à s'assurer que les erreurs de saisie des données soient réduites au maximum et enfin l'exécution de vérifications de la qualité du codage et du contrôle destinées à vérifier la logique du traitement. Malgré ces efforts, on ne peut éviter que des erreurs non dues à l'échantillonnage aient un certain effet sur les estimations de l'EPS. La section qui suit donne un aperçu des sources les plus probables de cette erreur et de ses effets sur les estimations de l'enquête.

9.1

Non-réponse complète

Dans beaucoup de sondages, la non-réponse complète peut être une source importante d'erreur, selon la mesure dans laquelle les répondants et les non-répondants diffèrent quant aux caractéristiques présentées. Dans l'EPS, il y a eu des cas de non-réponse complète lorsqu'un répondant n'a pu être rejoint ou a refusé de participer à l'enquête. L'analyse des caractéristiques des non-répondants donne à penser qu'ils ne sont pas concentrés dans un groupe donné. Pour remédier à la non-réponse complète, on réajuste le poids d'échantillonnage des répondants, compensant ainsi les données manquantes.

9.2 Non-réponse partielle

Dans l'EPS, la non-réponse partielle s'est produite lorsqu'un répondant a refusé de répondre à une question ou n'a pas compris une question. Lors du processus de vérification, on a déterminé les questions qui devaient avoir une réponse. Pour un enregistrement donné, si une de ces questions n'étaient pas répondu, on retranchait ledit enregistrement. 156 enregistrements ont été retranchés du fichier. Par conséquent, il est peu probable que la non-réponse partielle ait contribué de façon significative à l'erreur non due à l'échantillonnage.

9.3 Taux de réponse

Pour l'EPS, des 35 077 numéros de téléphone qui ont été signalés, 17 674 d'entre eux correspondaient à des ménages. De ces ménages, 2 288 (12.9 %) ont refusé de répondre ou n'ont pas pu répondre à l'enquête. Ce compte inclut aussi les ménages qui n'ont pas pu être rejoint tout au long de la période de collecte des données de l'enquête.

Des 15 386 ménages où on a tenté de mener une entrevue, 511 personnes choisies ont refusé de répondre au questionnaire (une personne par ménage était choisie au hasard). De plus, 168 enregistrements ont été retranchés du fichier lors du traitement des données au bureau chef soit à cause de la non-réponse partielle ou soit à cause de l'âge du répondant (moins de 15 ans).

Si on suppose que tous les ménages, qui n'ont pas répondu et qui ont été retranchés du fichier, faisaient parti de la population cible (c-à-d avaient au moins un membre âgé de 15 ans et plus), alors le taux de réponse global de l'enquête serait de 78.0 %.

9.4 Couverture

Tel que mentionné au chapitre 5.1, moins de 3 % de la population total n'a pas de téléphones. Des caractéristiques spécifiques pouvant être rattachées à cette parti de la population ne seront pas compris dans les estimés de l'enquête. Lors de l'interprétation d'estimés de cette parti de la population, les utilisateurs devraient tenir compte du biais qui pourrait exister lors de la corrélation de ces caractéristiques et le fait de ne pas posséder de téléphones.



9.5

Sous-estimation

Du fait de la nature délicate de certaines questions, notamment les questions sur la santé sexuelle, le taux de réponses varie selon que les questions sont plus ou moins personnelles (10 % de refus pour la question sur l'âge de la première relation sexuelle comparée à 2 % pour la question sur l'effet de la pollution sur la santé). Lors de l'interprétation des estimés, l'utilisateur devrait tenir compte du pourcentage de refus.



10.0

Lignes directrices relatives à la publication et à la diffusion

IL EST IMPORTANT QUE LES UTILISATEURS SE FAMILIARISENT AVEC LE CONTENU DE CETTE SECTION AVANT DE PUBLIER OU DE DIFFUSER TOUTE ESTIMATION PRODUITE A PARTIR DU FICHER DE MICRODONNÉES DE L'ENQUÊTE SUR LA PROMOTION DE LA SANTÉ.

Cette section du guide porte sur les règles que les utilisateurs doivent suivre pour la publication et la diffusion de toutes données tirées des fichiers de microdonnées de l'enquête. En suivant ces règles, les utilisateurs devraient pouvoir produire les mêmes chiffres que ceux qui sont produits par Statistique Canada et, en même temps, produire des données qui ne sont pas encore publiées, mais qui sont conformes aux règles établies. Cette section se divise essentiellement en quatre parties : les règles relatives à l'arrondissement, les règles relatives à la pondération de l'échantillon, les règles relatives à la variabilité d'échantillonnage et les règles relatives à l'analyse statistique.

10.1

Règles relatives à l'arrondissement

Afin que les estimations produites à partir de ce fichier de microdonnées et destinées à la publication ou à toute autre forme de diffusion correspondent à celles qui sont produites par Statistique Canada, les utilisateurs sont invités à suivre les règles suivantes concernant l'arrondissement des estimations.

- a) Dans le corps principal d'un tableau statistique, les estimations doivent être arrondies au millier le plus proche à l'aide de la technique d'arrondissement normale. Selon cette technique, si le premier ou le seul chiffre à retrancher se situe entre 0 et 4, le dernier chiffre à retenir demeure inchangé. Si le premier ou le seul chiffre à retrancher se situe entre 5 et 9, le dernier chiffre à retenir est augmenté d'une unité. Par exemple, pour l'arrondissement normal au millier près, si les trois derniers chiffres se situent entre 000 et 499, ils sont remis à 000 et le chiffre précédent (le chiffre des milliers) demeure inchangé. Si les derniers chiffres se situent entre 500 et 999, ils sont remis à 000 et on ajoute 1 au chiffre précédent.

- b) Dans les tableaux statistiques, les sous-totaux et totaux marginaux doivent être calculés à partir des chiffres non arrondis qui les composent et être ensuite arrondis eux-mêmes au millier près par la technique d'arrondissement normal.
- c) Les moyennes, les proportions, les taux et les pourcentages doivent être calculés à partir de chiffres non arrondis (c'est-à-dire les numérateurs et les dénominateurs) et être ensuite arrondis eux-mêmes à une décimale par la technique d'arrondissement normal. Lorsqu'on arrondit à un seul chiffre, si le premier ou le seul chiffre à retrancher se situe entre 0 et 4, le dernier chiffre à retenir demeure inchangé. Si le premier ou le seul chiffre à retrancher se situe entre 5 et 9, on ajoute 1 au chiffre à retenir.
- d) Les sommes des agrégats ou des rapports et les différences qui les séparent doivent être calculées à partir de chiffres non arrondis et être ensuite arrondies elles-mêmes au millier près ou à la décimale la plus près par la technique d'arrondissement normal.
- e) Si, à cause de limitations techniques ou autres, une méthode d'arrondissement autre que celle de l'arrondissement normal est utilisée, donnant ainsi lieu à des estimations à publier ou à diffuser qui diffèrent des estimations correspondantes publiées par Statistique Canada, les utilisateurs doivent indiquer la raison de ces différences au moment de la publication ou de la diffusion de leurs documents.
- f) Des estimations non arrondies ne peuvent en aucune circonstance être publiées ou diffusées par les utilisateurs. Les estimations non arrondies donnent l'impression que la précision est supérieure à ce qu'elle est en réalité.

10.2

Règles relatives à la pondération de l'échantillon dans les applications

Le plan de sondage de l'EPS n'est pas un plan auto-pondéré. Lorsqu'ils produisent des estimations simples, y compris des tableaux statistiques ordinaires, les utilisateurs doivent appliquer les poids d'échantillonnage qui figurent dans les enregistrements du fichier de microdonnées. Autrement, les estimations établies à partir de ce fichier ne pourront pas être considérées comme étant représentatives de la population observée et ne correspondront pas aux estimations produites par Statistique Canada.

Les utilisateurs doivent également noter qu'à cause de la configuration de leur zone de poids, certains logiciels ne permettent pas de produire des estimations qui correspondent exactement à celles qui sont produites par Statistique Canada.

10.3

Règles relatives à la variabilité d'échantillonnage

Les utilisateurs devraient d'abord déterminer le nombre de répondants sur le fichier microdonnées qui ont contribué au calcul des estimations. Si ce nombre est moins de 30, l'estimé (pondéré) ne doit pas être diffusé quelque soit la valeur du coefficient de variation de cette estimé (le coefficient de variation (cv) est le pourcentage de l'écart-type).

Avant la publication ou la diffusion d'un estimé tiré du fichier microdonnées, l'utilisateur doit déterminer le coefficient de variation approximatif de chaque estimé et suivre les lignes directrices ci-dessous. Le chapitre 8 explique en détails comment obtenir le coefficient de variation approximatif pour chaque estimé.

Avant de publier des estimations tirées des fichiers de microdonnées, les utilisateurs doivent déterminer leurs cv et suivre les règles suivantes.

Règles relatives à la variabilité d'échantillonnage pour L'Enquête sur la Promotion sur la Santé

| Type d'estimation | cv (en %) | Règles |
|-------------------|----------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1. Sans réserve | 0.0 - 16.5 % | Les estimations peuvent faire l'objet d'une diffusion générale sans restriction. Aucune mention spéciale n'est nécessaire. |
| 2. Avec réserve | 16.6 - 25.0 % | Les estimations peuvent faire l'objet d'une diffusion générale sans restriction, mais elles doivent être accompagnées d'une note avertissant les utilisateurs de leur variabilité d'échantillonnage élevée. Ces estimations doivent être identifiées par la lettre G (ou par une mention équivalente). |
| 3. Confidentiel | 25.1 - 33.3 % | Les estimations peuvent faire l'objet d'une diffusion générale seulement si la variabilité d'échantillonnage est obtenue par une méthode de calcul exact de la variance. À moins de comporter ces variances, ces estimations doivent être éliminées et remplacées par des tirets (---) dans les tableaux statistiques. |
| 4. Ne pas publier | 33.4 % ou plus | Les estimations ne doivent être publiées sous aucune forme et à aucune condition. Dans les tableaux statistiques, ces estimations doivent être éliminées. |

Remarque : Ces règles relatives à la variabilité d'échantillonnage doivent être appliquées à des estimations arrondies.

10.4

Règles relatives à l'analyse statistique

L'EPS a un plan de sondage complexe comportant de la stratification, plusieurs degrés d'échantillonnage et des probabilités inégales de sélection des répondants. L'utilisation de données d'enquêtes aussi complexes pose des problèmes aux analystes, car le plan du sondage et les probabilités de sélection ont une influence sur les méthodes d'estimation et de calcul de la variance qui doivent être utilisées.

Même si de nombreuses procédures d'analyse se trouvant dans des progiciels statistiques permettent de faire de la pondération, le sens ou la définition qu'on donne aux poids dans ces procédures diffère de celui qui est approprié dans le contexte d'une enquête par sondage; il en résulte que, même si les estimations produites par les progiciels sont correctes dans de nombreux cas, les variances calculées n'ont presque aucune signification.

Il existe une méthode qu'on peut appliquer à de nombreuses techniques d'analyse (par exemple la régression linéaire, la régression logistique, l'évaluation des taux et des proportions et l'analyse des variances) et qui permet de rendre les variance calculées au moyen de progiciels standard plus significatives. Si les poids indiqués dans les enregistrements sont recalculés de façon à ce que le poids moyen soit un (1), les variances produites par les progiciels standard seront alors plus acceptables; elles ne tiendront toujours pas compte de la stratification et de la formation de grappes prévues par le plan de sondage, mais elles tiendront compte des probabilités inégales de sélection. Pour recalculer les poids, il suffit de diviser chaque poids par le poids moyen global avant d'entreprendre l'analyse.



11.0 Questionnaires





12.0

Méthodes de pondération

12.1

Estimation

Lorsqu'on utilise un échantillon probabiliste, comme dans le cas de l'EPS, le principe qui sous-tend l'estimation est que chaque personne choisie pour l'échantillon représente, outre elle-même, plusieurs autres personnes qui ne font pas partie de l'échantillon. Par exemple, dans un échantillon aléatoire simple représentant 1 % de la population, chaque personne de l'échantillon remplace 100 personnes de cette population.

Pour chaque fichier de microdonnées, on attribue un poids statistique global (que l'on nomme "POIDS") à chaque enregistrement afin de tenir compte du nombre de personnes échantillonnées représentées par l'enregistrement. Ce facteur de pondération fait référence au nombre de fois qu'un enregistrement particulier doit contribuer à l'estimation de la population. Par exemple, afin d'estimer le nombre de personnes qui décrivent leur vie comme "très stressante", la valeur du poids est additionnée pour tous les enregistrements où la question A2 a un code égal à 1. La méthode de pondération de l'EPS est décrite à la section 12.2 ci-dessous.

12.2

Pondération de l'Enquête Promotion Santé

Comme on a utilisé deux techniques d'échantillonnage (nous l'avons vu à la section 5) pour l'EPS, il a fallu se servir de deux méthodes de pondération légèrement différentes. Les méthodes de pondération préliminaires pour chacun des plans de l'EPS sont décrits séparément, en fonction de la technique d'échantillonnage utilisée. Les rajustements finals apportés à tous les enregistrements sont décrits dans une troisième section.

12.2.1

Méthode de pondération préliminaire pour le plan de Waksberg

La méthode de Waksberg a servi pour la strate de l'Île-du-Prince-Édouard et la strate du Québec comprenant les régions qui ne sont pas des régions métropolitaines de recensement.

Un plan de sondage à autopondération est un plan dans lequel le poids attribué à toutes les unités de l'échantillon est le même. Dans le cas d'un plan de sondage à deux degrés, il y a autopondération lorsque les unités du premier degré (c.-à-d. les unités primaires d'échantillonnage) sont choisies suivant une probabilité proportionnelle à l'échantillon et qu'un nombre déterminé d'unités est choisi suivant une même probabilité dans chaque unité primaire d'échantillonnage prélevée.

Voici dans leurs grandes lignes, les étapes suivies pour la pondération des enregistrements de l'EPS au moyen de la méthode de Waksberg.

1) Poids de base

La première étape de la pondération a consisté à attribuer un poids égal à tous les ménages qui avaient été échantillonnés dans une strate donnée.

2) Rajustement pour la non-réponse

Les poids des ménages répondants ont été rajustés afin de représenter les ménages non répondants. Selon le plan de sondage de l'EPS, il fallait rejoindre six ménages pour chaque banque active de numéros de téléphone échantillonnés. Dans certains cas, un ou plusieurs de ces six ménages refusaient de participer à l'enquête. Les poids des ménages répondants étaient alors rajustés pour tenir compte des ménages non répondants : le poids de base des ménages répondants de la banque était multiplié par le rapport suivant :

$$\frac{6}{\text{Nombre de ménages répondants dans la banque}}$$

3) Rajustement pour tenir compte des ménages ayant plusieurs numéros de téléphone

On a diminué les poids pour les ménages ayant plusieurs numéros de téléphone privés afin de tenir compte du fait qu'il existe une plus forte probabilité qu'ils soient choisis. Le poids de chaque ménage a ainsi été divisé par le nombre de numéros de téléphone distincts qui desservent ce ménage.

4) Calcul du poids des personnes

On a ensuite calculé un poids pour chaque personne répondant à l'enquête en multipliant le poids du ménage auquel elle appartenait par le nombre de membres du ménage susceptibles d'être choisis pour l'enquête (soit le nombre de membres du ménage âgés de 15 ans ou plus).

12.2.2

Méthode de pondération préliminaire pour le plan de l'élimination des banques inactives

Comme pour la méthode de Waksberg, tous les ménages d'une strate donnée ont la même probabilité de sélection avec la méthode d'élimination des banques inactives (MEBI). Cette probabilité est égale à :

$$\frac{\text{Nombre de numéros de téléphone échantillonnés dans la strate}}{\text{Nombre total de numéros de téléphone possibles dans la strate}}$$

Il convient de remarquer que le nombre total de numéros de téléphone possibles dans une strate est égal au nombre de banques actives dans une strate multiplié par cent.

Les étapes ci-dessous décrivent dans ses grandes lignes la méthode de pondération utilisée pour les enregistrements de la MEBI.

1) Poids de base

On a attribué à chaque ménage (répondant et non répondant) un poids égal à l'inverse de sa probabilité de sélection, soit :

$$\frac{\text{Nombre total de numéros de téléphone possibles dans la strate}}{\text{Nombre de numéros de téléphone échantillonnés dans la strate}}$$

2) Rajustement pour la non-réponse

Les poids des ménages répondants ont été rajustés afin de représenter les ménages non répondants. Ce rajustement s'est effectué indépendamment pour chaque indicatif régional. Les enregistrements ont été rajustés en fonction du facteur suivant :

$$\frac{\text{Somme des poids des ménages pour tous les ménages d'un indicatif régional}}{\text{Somme des poids des ménages pour les ménages répondants d'un indicatif régional}}$$

Les ménages non répondants ont ensuite été éliminés des pondérations subséquentes.

3) Rajustement pour tenir compte des ménages ayant plusieurs numéros de téléphone

On a abaissé les poids pour les ménages ayant plusieurs numéros de téléphone privés afin de tenir compte du fait qu'il existe une plus forte probabilité qu'ils soient choisis. Le poids de chaque ménage a ainsi été divisé par le nombre de numéros de téléphone distincts qui desservent ce ménage.

4) Calcul du poids des personnes

On a ensuite calculé un poids pour chaque personne répondant à l'enquête en multipliant le poids du ménage auquel elle appartenait par le nombre de membres du ménage susceptibles d'être choisis pour l'enquête (soit le nombre de membres du ménage âgés de 15 ans ou plus).

12.2.3

Méthode de pondération combinée pour les deux plans

Une fois appliquées les méthodes de pondération préliminaires pour le plan de Waksberg et le plan d'élimination des banques inactives, les fichiers séparés ont été réunis pour la méthode de pondération "combinée".

5) Rajustement pour les totaux externes à la strate

On a procédé à un rajustement des poids des personnes pour les enregistrements de chaque strate afin d'établir des estimations démographiques qui correspondent aux estimations démographiques du recensement. On a effectué ce rajustement en multipliant le poids des personnes de chaque enregistrement dans la strate par le rapport suivant :

$$\frac{\text{Projection démographique du recensement pour les personnes âgées de 15 ans et plus de la strate}}{\text{Somme des poids des personnes des enregistrements du plan de Waksberg de l'EPS dans la strate}}$$

6) Rajustement pour les totaux de la province, du sexe et des groupes d'âge

L'étape suivante de la pondération a consisté à rajuster suivant un rapport les poids de tous les enregistrements d'une province afin qu'ils correspondent aux répartitions âge-sexe du recensement projetées. Le mois de référence des projections démographiques du recensement est juin 1990.

Les groupes d'âge suivants ont été établis tant pour les hommes que pour les femmes :

15-19, 20-24, 25-29, 30-34, 35-39, 40-44,
45-49, 50-54, 55-59, 60-64, 65-69, 70 et plus

Pour chacune des catégories établies (10 provinces x 12 groupes d'âge x 2 sexes), on a rajusté les poids des personnes pour les enregistrements de la classification en les multipliant par le rapport suivant :

$$\frac{\text{Population du recensement projetés} \\ \text{- sexe - groupe d'âge}}{\text{Somme des poids des personnes pour les enregistrements} \\ \text{de la province - sexe - groupe d'âge}}$$

Il a fallu réunir certains groupes d'âge avant d'utiliser ce rapport afin de respecter les critères minimaux de taille.

Il faut remarquer que les personnes appartenant à un ménage sans service téléphonique ont été intégrées à ces projections même si elles n'ont pas été échantillonnées.

7) Rajustement par la méthode itérative du quotient


Les poids de chaque répondant ont été rajustés plusieurs fois au moyen de la méthode itérative du quotient. Cette méthode permet de faire en sorte que les estimations produites pour une strate et pour chaque province - sexe - groupe d'âge correspondent.

On a effectué ce rajustement en répétant les étapes 5) et 6) des méthodes de pondération au moyen de poids obtenus à l'étape précédente, jusqu'à ce que les deux séries d'estimations soient correctes. Le poids statistique final est devenu la variable "POIDS".

12.3 Politique concernant la pondération

On conseille fortement aux utilisateurs de ne pas diffuser des tableaux non pondérés ou des analyses fondées sur les résultats non pondérés de l'enquête car plusieurs rajustements de pondération ont été accomplis indépendamment sur les enregistrements de chaque province. De plus, l'échantillonnage et le taux de réponse variaient grandement d'une province à l'autre et d'un groupe d'âge à l'autre.

L'EPS a été conçue de manière à ce qu'il soit possible de se fier aux estimations produites au niveau provincial. En raison des différences de population dans bon nombre de strates, il existe une importante différence entre les poids moyens les plus faibles et les plus élevés de l'enquête. Par exemple, on a attribué un poids moyen d'environ 105 aux répondants de l'Î.-P.-É. alors que le poids moyen des répondants du Québec était supérieur à 300.



Par ailleurs, on sait qu'en général les non-répondants sont d'une part de sexe masculin et, d'autre part, jeunes. Dans l'échantillon de l'EPS, les hommes âgés de 15 à 19 ans représentaient 3.6 % de l'échantillon brut mais 4.6 % de la population pondérée. Par contre, les femmes âgées de 70 ans ou plus constituaient 4.0 % de l'échantillon brut mais seulement 2.8 % de la population.

De toute évidence, les chiffres de l'échantillon ne peuvent être considérés comme représentatifs de la population visée par l'enquête à moins que les poids appropriés n'aient été appliqués.



13.0

Cliché d'articles





14.0

Notes explicatives sur le cliché d'articles

Notes relatives à l'interprétation du cliché d'articles de l'enquête sur la promotion de la santé.

Ces notes visent à donner des renseignements supplémentaires concernant des zones qui ne sont peut-être pas clairement définies sur le cliché d'articles. Il faut donc les lire en même temps que celui-ci. On avertit les utilisateurs que, dans certains cas, les numéros de case sur le questionnaire ne sont pas les mêmes que ceux qui figurent sur le cliché. Il faut donc toujours consulter le cliché d'articles pour comprendre les variables dans le fichier.

14.1

Sigle

Le sigle utilisé sur le cliché d'articles se rapporte au numéro de la question tel qu'il apparaît sur le questionnaire. Exception faite pour les variables dérivées et les questions à choix multiples. Par exemple, la question C1, une question à choix multiples, sera ainsi représenté sur le cliché d'articles :

- C1CAT01: où, C1 est le numéro de la question, CAT la catégorie, et 01 le numéro de la catégorie.

N.B.: Lors de l'entrevue, les catégories de réponse des questions utilisant ce format (C1, C5, C8, I2, N4, Q2, Q4, R4 and R5) n'ont pas été lues, cependant l'intervieweur a coché la catégorie correspondant à la réponse donnée.

14.2

Codification

Tout au long du questionnaire, une structure de codification normalisée a été utilisée dans la mesure du possible.

Code 8, 98, 998 :

Ce code veut dire que le répondant n'avait à répondre à la question, elle n'a pas été posée.

Du fait que la section L s'applique aux répondants de sexe féminin seulement, ce code a été utilisé pour toutes les questions de cette section si le répondant était de sexe masculin. La même logique a été appliquée à la section M pour les répondants de sexe féminin.

Code 9, 99, 999 :

Ce code veut dire que le répondant n'a pas répondu à la question. Cependant il aurait du répondre.

Pour les questions à choix multiples, si le répondant n'a pas répondu à toutes les catégories, alors toutes les catégories auront ce code. Si une catégorie a été répondue, alors le code de la catégorie répondue sera 1, et les autres seront 2.

14.3 Codes (valeurs)

Les codes du cliché d'articles ne correspondent pas toujours aux codes imprimés sur le questionnaire. De plus, les explications des «notes» sur le cliché d'articles se rapportent aux codes du cliché et non aux codes du questionnaire. Par exemple, la «note» de la question E3a explique que la question n'est pas demandée si E1=1 ou E2=2. Si on se réfère au questionnaire, les codes utilisés pour E1 sont valides mais non ceux de E2 (la valeur correspondante de 2 est 4 sur le questionnaire).



15.0

Tableaux de la variabilité d'échantillonnage approximative

