

Chapitre 2 - Qualité des données

Le chapitre 2 fournit des notes sur la qualité des données que procure le fichier. Il comprend deux parties:

Renseignements relatifs aux propriétés de l'échantillon

Le lecteur trouvera ici les renseignements concernant le plan de sondage et la fiabilité des estimations, c'est-à-dire:

- A. L'univers visé et les limitations géographiques du fichier
- B. Le plan de sondage
- C. Les estimations
 - 1. Pondération
 - 2. Genres d'estimations
- D. La fiabilité des données
 - 1. Concordance des estimations
 - 2. Variabilité d'échantillonnage
 - 3. Variabilité non due à l'échantillonnage

Renseignements relatifs aux autres facteurs qui affectent la fiabilité des données

Des erreurs non dues à l'échantillonnage peuvent également influencer la qualité des données. L'utilisateur trouvera ici des renseignements sur le soutien du ménage et le type de logement, deux variables dont il faut interpréter les données avec prudence.

PLAN DE SONDAGE ET ESTIMATION

A. Univers visé et limitations géographiques du fichier

La population visée par le fichier comprend tous les ménages privés dans les logements privés occupés hors réserve au Canada. Ainsi, les sous-groupes de population suivants sont exclus du fichier:

- a) les ménages collectifs;
- b) les ménages à l'extérieur du Canada;
- c) les logements occupés par des résidents temporaires et (ou) étrangers;
- d) les ménages situés sur des réserves indiennes.

Dans le but de satisfaire des critères de confidentialité, l'information géographique contenue dans le fichier est limitée. Ainsi, la région métropolitaine de recensement (RMR) ventilée par le code rural/urbain offre le plus petit niveau d'agrégation géographique du fichier. Toutefois, un critère supplémentaire est imposé, soit celui d'une population minimale par RMR. Ainsi, seules les RMR avec une population de ménages privés hors réserve de 50,000 et plus sont identifiées. Pour cette raison, certaines RMR sont combinées, soit:

Trois-Rivières et Sherbrooke
Sudbury et Thunder Bay

On fait toutefois deux exceptions à la règle de population minimale: la province de l'Île-du-Prince-Édouard et les Territoires du Nord-Ouest/Yukon sont identifiés séparément.

Pour une liste complète des RMR identifiées, le lecteur se référera au chapitre 1.

B. Plan de sondage

L'échantillon de microdonnées des ménages a été choisi par une méthode d'échantillonnage à deux degrés. Le premier degré d'échantillonnage est constitué de l'échantillon d'un cinquième (données 2B) recueilli lors du recensement de 1986. Pour le deuxième degré d'échantillonnage, on a sélectionné de façon systématique, à l'intérieur de strates préalablement définies, un échantillon de ménages avec probabilité proportionnelle au poids du ménage.

1. Premier degré d'échantillonnage

Au recensement de 1986, quatre ménages sur cinq ont été dénombrés à l'aide d'un questionnaire abrégé (2A). Sur ce questionnaire, on retrouvait neuf questions de nature démographique et ethno-culturelle/langue. Un ménage sur cinq a, quant à lui, reçu un questionnaire complet (2B). En plus des neuf questions du questionnaire 2A, le questionnaire 2B comportait 23 autres questions définissant un large éventail de sujets.

Le premier degré d'échantillonnage du fichier de microdonnées des ménages considère donc l'échantillon d'un cinquième (2B) du recensement. Les données recueillies au recensement pour ce sous-groupe de la population sont pondérées; ainsi, chaque ménage dans l'échantillon "représente", outre lui-même, quatre ménages qui ne font pas partie de l'échantillon. Le coefficient de pondération ou poids de chaque ménage se situe en moyenne autour de cinq; toutefois, une procédure statistique d'ajustement de ces poids les modifie quelque peu. Pour plus de détails concernant l'échantillon d'un cinquième du recensement, veuillez consulter "Le recensement en bref", 1988, n° 99-104F au catalogue.

2. Deuxième degré d'échantillonnage

La population visée par le fichier, telle que mentionnée à la section A, a été divisée en sous-groupes ou strates géographiques et un échantillon de ménages a été tiré de chacune de ces strates. On a sélectionné en moyenne 1.29 ménage à tous les 100 ménages, soit environ 6.45 ménages à tous les 100 ménages de la population 2B non pondérée. Cette fraction de sondage a varié selon la région géographique pour permettre, à part quelques exceptions, l'obtention d'une taille d'échantillon minimale de 2,500 ménages. On présente au tableau 1 les strates géographiques accompagnées de leur fraction de sondage inverse, de la taille de la population visée et de la taille de l'échantillon obtenu.

Tableau 1: Stratification géographique de l'échantillon de ménages

Région géographique	Fraction de sondage inverse	Taille de la population visée (2B pondérée)	Taille de l'échantillon
<u>RMR</u>			
1. Halifax	41	103,828	2,532
2. Chicoutimi - Jonquière	20	51,274	2,564
3. Montréal	100	1,115,357	11,154
4. Ville de Québec	87	218,050	2,506
5. Trois-Rivières + Sherbrooke	38	95,992	2,526
6. Hull	28	70,274	2,510
7. Ottawa	92	232,002	2,522
8. Hamilton	80	201,327	2,516
9. Kitchener	44	110,154	2,504
10. London	51	129,402	2,537
11. Oshawa	27	68,010	2,519
12. St. Catharines-Niagara	49	124,576	2,543
13. Sudbury + Thunder Bay	38	95,089	2,503
14. Toronto	100	1,199,754	11,998
15. Windsor	36	91,613	2,545
16. Winnipeg	94	236,319	2,514
17. Regina	27	67,641	2,505
18. Saskatoon	29	73,920	2,549
19. Calgary	99	248,588	2,511
20. Edmonton	100	283,146	2,832
21. Vancouver	100	530,310	5,303
22. Victoria	41	104,654	2,553
<u>Régions résiduelles</u>			
23. Terre-Neuve	63	159,081	2,525
24. Nouveau-Brunswick	92	230,807	2,509
25. Île-du-Prince-Édouard	20	40,594	2,029
26. Nouvelle-Écosse (moins RMR)	100	190,813	1,908
27. Québec (moins RMR)	100	801,583	8,016
28. Ontario (moins RMR)	100	962,317	9,623
29. Prairies (moins RMR)	100	651,719	6,517
30. Colombie-Britannique (moins RMR)	100	439,675	4,396
31. Territoires du Nord-Ouest + Yukon	20	21,556	1,078
TOTAL		8,949,425	115,347

Ainsi, on note qu'à Halifax un ménage à tous les 41 ménages, soit 5 ménages sur 41 de la population 2B non pondérée, a été choisi. De plus, pour assurer une meilleure représentativité de l'échantillon, un tri de la population visée a été effectué à l'intérieur de chaque strate géographique selon les variables suivantes:

a) Mode d'occupation

- (i) propriétaire
- (ii) locataire

b) Type du ménage

- (i) ménage unifamilial
- (ii) ménage multifamilial
- (iii) ménage non familial

c) Revenu du ménage

- (i) moins de \$24,000
- (ii) \$24,000 et plus

Pour les ménages situés dans des RMR, un tri supplémentaire est effectué par:

d1) Région de résidence

- (i) noyau urbain
- (ii) banlieue urbaine
- (iii) banlieue rurale

Pour les ménages des régions résiduelles, le dernier tri est:

d2) Région de résidence

- (i) population de 100,000 et plus
- (ii) population de 30,000 à 99,999
- (iii) population de 10,000 à 29,999
- (iv) population de 2,500 à 9,999
- (v) population de moins de 2,500 (incluant les régions rurales)

L'échantillon a ensuite été sélectionné à l'intérieur de chacune des strates de la façon suivante: on a choisi un nombre aléatoire appelé "départ aléatoire" situé entre 0 et la fraction de sondage inverse moins 1. On a additionné au départ aléatoire le poids du premier ménage de la strate tel que défini par l'échantillon 2B. Si la somme obtenue était plus grande ou égale à la fraction de sondage inverse, le ménage était choisi; sinon, on additionnait le poids du ménage suivant et ainsi de suite. On appelle cette procédure "échantillonnage systématique avec probabilité proportionnelle à la taille", la taille étant ici représentée par le poids du ménage. Cette méthode permet d'obtenir un échantillon pour lequel la répartition des caractéristiques des ménages est la même que celle observée dans la population totale. De manière à éviter tout risque d'identification des ménages due au caractère systématique de l'échantillonnage, l'ordre des ménages a été modifié de façon aléatoire à l'intérieur de chaque strate géographique.

C. Estimations

Dans cette section, on introduit la notion de pondération dans le contexte du fichier de microdonnées des ménages, puis on explique brièvement les genres d'estimations pouvant être obtenus à l'aide du fichier de microdonnées.

1. Pondération

Le fichier de microdonnées des ménages contient un enregistrement pour chaque ménage sélectionné dans l'échantillon. Chaque enregistrement renferme un certain nombre de caractéristiques ou variables décrites au chapitre 1. Or, chacun de ces ménages représente, outre lui-même, plusieurs autres ménages qui ne font pas partie de l'échantillon. Au lieu de répéter les enregistrements de l'échantillon suivant le nombre de ménages qu'ils représentent, on attribue à chaque enregistrement un coefficient de pondération ou poids. Ce poids est ici équivalent à la fraction de sondage inverse associée à la région géographique particulière au ménage. Ces fractions inverses sont énumérées au tableau 1 et identifiées par la variable WEIGHT dans le fichier de microdonnées.

Le poids indique donc le nombre de fois qu'un enregistrement particulier doit être reproduit pour obtenir des estimations de la population. Par exemple, si l'on veut estimer le nombre de logements "semi-détachés" au Canada, on choisira les enregistrements du fichier de microdonnées se rapportant à cette catégorie de logement, puis on fera la somme des poids attribués à ces enregistrements.

Attention: Les utilisateurs doivent s'assurer de ne pas diffuser des tableaux non pondérés ni de faire d'analyses fondées sur des données non pondérées du fichier de microdonnées. En effet, le fichier est obtenu d'un plan de sondage complexe, comportant des différences importantes dans les fractions de sondage entre strates géographiques (tableau 1). Par conséquent, certaines régions sont surreprésentées dans l'échantillon par rapport à leur population, pour permettre l'obtention d'une taille d'échantillon minimale de 2,500 ménages à l'intérieur de chaque strate géographique. Cela signifie donc que l'échantillon non pondéré n'est pas représentatif de la population visée par le fichier.

2. Genres d'estimations

Deux types de variables sont contenues dans le fichier: des variables qualitatives indiquant la présence ou l'absence pour un ménage donné d'une caractéristique particulière, et des variables quantitatives tel le revenu.

a) Estimation dans le cas de variables qualitatives

Supposons qu'on veuille estimer le nombre de soutiens du ménage résidant en Ontario et d'origine ethnique britannique. On obtient facilement cette estimation en faisant la somme des poids des enregistrements

correspondant aux ménages possédant cette combinaison de caractéristiques. On peut également exprimer ce total sous la forme d'une proportion de tous ces soutiens du ménage résidant en Ontario. Ce dernier total est tout simplement obtenu en faisant la somme des poids de tous les ménages de la province.

b) Estimation dans le cas de variables quantitatives

Dans le cas de variables quantitatives, on peut obtenir des estimations basées directement sur les valeurs numériques de celles-ci, tels la moyenne, la médiane ou le total. Par exemple, on pourrait calculer le revenu d'emploi moyen des soutiens du ménage de sexe féminin au Québec. On considère ainsi le rapport suivant:

$$\begin{array}{l} \text{Estimation du revenu d'emploi} \\ \text{moyen des soutiens du} \\ \text{ménage de sexe féminin au Québec} \end{array} = \frac{X}{Y}$$

où le numérateur (X) est une estimation quantitative du revenu d'emploi total des soutiens du ménage de sexe féminin au Québec. Cette estimation s'obtient en multipliant le poids de chacun des enregistrements (un par ménage) des soutiens du ménage de sexe féminin au Québec par leur revenu d'emploi, et en faisant la somme de ces produits pour tous les enregistrements appropriés. Le dénominateur (Y) s'obtient en faisant la somme des poids des enregistrements associés aux soutiens du ménage de sexe féminin au Québec. On note que les estimations X et Y sont obtenues de façon indépendante puis divisées l'une par l'autre.

c) Analyses plus complexes

L'utilisateur peut vouloir appliquer certaines techniques d'analyses de données au fichier sur les ménages, telles une régression linéaire, logistique ou une analyse de la variance. Il doit toutefois être très prudent. En effet, le fichier de microdonnées est obtenu d'un plan de sondage complexe et à chaque enregistrement est donc associé un poids. L'utilisateur doit tenir compte de ces poids lors de son analyse. L'analyse particulière appliquée aux données sera significative dans la mesure où elle tient compte du plan d'échantillonnage.

D. Fiabilité des données

Puisque l'échantillon de microdonnées est basé sur un échantillon de ménages, on ne peut pas s'attendre à une concordance parfaite entre les données du recensement (basées sur le recensement complet ou sur l'échantillon d'un cinquième du recensement) et les estimations fondées sur les résultats obtenus à l'aide des enregistrements de microdonnées. Ces données différeront inévitablement dans une certaine mesure, en raison du caractère aléatoire du choix de l'échantillon. Or, on peut estimer ce degré de variation dû au hasard.

Des tableaux généraux montrant des mesures de l'erreur d'échantillonnage sont présentés ci-après. On traite toutefois auparavant d'un test statistique visant à évaluer le degré de concordance de l'échantillon de microdonnées par rapport aux données 2B du recensement. On introduit finalement une brève note concernant les erreurs non dues à l'échantillonnage.

1. Concordance des estimations

Un test statistique souvent utilisé pour vérifier le degré de concordance des estimations faites à partir d'un échantillon et des totaux de l'ensemble de la population est le test du chi-deux. On ne présente pas, dans le cadre du présent document, la description mathématique de ce test. Notons toutefois qu'il s'intéresse, pour chaque variable du fichier, à la distribution observée dans l'échantillon par rapport à la distribution obtenue de la population du recensement (pondérée 2B) et ce, à l'intérieur de chaque région géographique incluse dans le fichier. Un niveau critique est fixé, en l'occurrence 5%, et on peut s'attendre qu'en moyenne 1 test sur 20 aura des résultats significatifs dus à la variabilité d'échantillonnage. Si on excède cette limite, l'échantillon de microdonnées constitue un mauvais échantillon de la population du recensement pour une variable donnée.

Les résultats ont été conformes aux prévisions puisque moins de 5% des tests statistiques calculés étaient significatifs. On peut donc supposer que l'échantillon obtenu représente de manière acceptable la population du recensement.

2. Variabilité d'échantillonnage

La différence entre les estimations obtenues de l'échantillon de microdonnées des ménages et les résultats du recensement est définie comme l'erreur d'échantillonnage de l'estimation. Or, même si on ne connaît pas la valeur exacte de cette différence, il est possible d'estimer une mesure statistique appelée erreur type à partir des données de l'échantillon. Cette mesure reflète la variabilité à laquelle on pourrait s'attendre pour l'estimation d'une caractéristique si plusieurs échantillons du même type étaient tirés.

À l'aide de l'erreur type, il est possible de définir des intervalles de confiance pour les estimations en supposant qu'elles sont réparties de façon normale autour de la valeur réelle de la population. Ainsi, il y a environ 68% des chances que la différence entre une estimation faite à partir de l'échantillon et la valeur réelle de la population soit inférieure à une erreur type, environ 95% des chances que cette différence soit inférieure à deux erreurs types et environ 99% des chances qu'elle soit inférieure à deux fois et demie l'erreur type.

Il est à noter que ces intervalles ne tiennent compte que de la variabilité d'échantillonnage de l'estimation. Ainsi, les erreurs non dues à la variabilité d'échantillonnage, telles les erreurs de réponse et de traitement qui peuvent introduire des biais systématiques dans les données, ne sont pas considérées par l'intervalle de confiance. On discute des erreurs non dues à l'échantillonnage à la section D.3.

a) Variabilité d'échantillonnage pour des variables qualitatives

Une mesure souvent utilisée pour montrer le degré de variabilité d'échantillonnage d'une estimation est le "coefficient de variation (CV)". Celui-ci est tout simplement le rapport, exprimé en pourcentage, de l'erreur type d'une estimation à la valeur de cette estimation.

On propose à l'annexe B des tableaux généraux de la variabilité d'échantillonnage. On note toutefois, qu'étant donné le grand nombre d'estimations pouvant être obtenu du fichier de microdonnées des ménages, il est impossible de présenter ici des CV **exacts** pour tous les domaines d'études possibles. Ainsi, des CV approximatifs en pourcentage sont présentés au niveau du Canada et pour chaque RMR et province identifiées dans le fichier de microdonnées. En plus d'être facile d'utilisation, ces tableaux, quoique approximatifs, permettent à l'utilisateur de décider si une estimation particulière peut faire l'objet d'une diffusion générale ou si, au contraire, elle doit être complètement ignorée.

Le tableau suivant indique les règles généralement utilisées à Statistique Canada pour décider si une estimation peut être diffusée ou non en fonction de son CV.

Catégorie	Coefficient de variation (%)	Code alphabétique	Recommandation
1. Sans restriction	0.0 à 0.5	A	Les estimations peuvent faire l'objet d'une diffusion générale sans restriction. L'utilisation du code alphabétique est recommandée. La lettre A indique qu'il s'agit d'une estimation très fiable; la lettre B, que l'estimation est fiable mais à un degré moindre qu'une estimation de la catégorie A, et ainsi de suite.
	0.6 à 1.0	B	
	1.1 à 2.5	C	
	2.6 à 5.0	D	
	5.1 à 10.0	E	
	10.1 à 16.5	F	
2. Avec restriction	16.6 à 25.0	G	Les estimations sont suffisamment fiables pour des utilisations particulières, mais il faut s'en servir avec beaucoup de circonspection. Chaque fois qu'on y aura recours, il faudra préciser que leur variabilité d'échantillonnage est plus élevée.
	25.1 à 33.3	H	
3. Ne pas diffuser	33.4 et plus	I	Les estimations ne doivent être diffusées sous aucune forme et sous aucune condition. Elles doivent être supprimées des tableaux statistiques.

Les CV présentés dans les Tableaux de la variabilité d'échantillonnage approximative (annexe B) ont été calculés pour un échantillon aléatoire simple. Toutefois, puisque l'échantillon de microdonnées des ménages est sélectionné suivant un plan d'échantillonnage plus complexe, on a multiplié chaque CV par un facteur appelé effet du plan de sondage. Ce facteur se définit comme le rapport entre l'erreur type réelle de l'estimation (selon le plan de sondage complexe) et l'erreur type qui découlerait de l'utilisation d'un échantillon aléatoire simple de même taille. On a calculé, pour chacune des RMR et chaque province et pour l'ensemble du Canada, l'effet du plan de sondage pour plusieurs caractéristiques du fichier. Les effets du plan de sondage utilisés dans les Tableaux de la variabilité d'échantillonnage approximative (annexe B) ont été choisis parmi l'ensemble des valeurs calculées. On a favorisé le choix de facteurs élevés plutôt que faibles introduisant ainsi un certain conservatisme dans les tableaux en montrant des CV en général plus grands que ceux qui auraient été obtenus de techniques plus précises.

On présente ci-après quelques règles d'utilisation des tableaux qui devraient permettre aux utilisateurs de déterminer les CV approximatifs pour des estimations de totaux, de pourcentages et de ratios.

Règle 1. Estimation de totaux

Dans ce cas, le CV dépend uniquement de la valeur du total estimé. Par exemple, si on estime à 532,000 le nombre de logements "semi-détachés" à Montréal, en utilisant le tableau de Montréal, on se rapporte à la rangée la plus près de 532,000 (numérateur du pourcentage) soit 500,000, puis on se déplace de façon horizontale jusqu'à ce qu'on rencontre une colonne possédant une estimation (colonne différente de "*****"). Pour cet exemple le CV approximatif se rapportant au nombre estimé de "semi-détaché" à Montréal se situe autour de 1.2%. On peut donc diffuser sans restriction cette estimation, l'utilisation du code alphabétique "C" étant toutefois recommandée.

Règle 2. Estimation de pourcentages

Le CV d'une estimation de pourcentage dépend de la valeur du pourcentage et de la taille de la population sur laquelle ce pourcentage est fondé. Par exemple, supposons qu'on désire obtenir le CV de l'estimation du nombre de soutiens du ménage dont l'âge se situe entre 20 et 24 ans, qui résident à Winnipeg et qui sont nés au Manitoba. L'estimation du nombre de soutiens du ménage dont l'âge se situe entre 20 et 24 ans à Winnipeg est de 17,578. De ce nombre, 12,690 ont comme lieu de naissance le Manitoba. Le numérateur du pourcentage est donc ici de 12,690. Le pourcentage estimé est de $12,690/17,578$, soit environ 72%. En utilisant le tableau de Winnipeg, on se rapporte à la rangée la plus près de 12,690, soit 12,000, puis à la colonne la plus près de 72%, soit 70%. À l'intersection de cette rangée et de cette colonne, on trouve le CV approximatif, soit 5.7%. L'estimation peut donc être diffusée sans restriction. L'utilisation du code alphabétique "E" est toutefois recommandée.

Règle 3. Estimation de ratios

Lorsque le numérateur est un sous-ensemble du dénominateur, comme dans l'exemple précédent, la règle 2 doit être utilisée. Lorsque le numérateur n'est pas un sous-ensemble du dénominateur comme, par exemple, le ratio du nombre de soutiens du ménage de sexe féminin au nombre de soutiens du ménage de sexe masculin, le CV est obtenu de la façon suivante:

$$\text{soit le ratio} \quad R = \frac{X}{Y}$$

$$CV(R) = (CV(X)^2 + CV(Y)^2)^{1/2}$$

où CV(X) et CV(Y) sont obtenus de la règle 1 ou 2. Toutefois, cette formule aura pour effet de surestimer le CV s'il existe une corrélation positive entre X et Y et de le sous-estimer s'il existe une corrélation négative.

b) Variabilité d'échantillonnage pour des variables quantitatives

Comme on l'a expliqué à la section C.2., il est possible de dériver, pour les variables quantitatives, des statistiques fondées sur la valeur numérique de ces variables. Les moyennes, les médianes et les coefficients de corrélation, par exemple, se rangent dans cette catégorie.

Toutefois, à chaque enregistrement du fichier de microdonnées est associé un poids et celui-ci doit être pris en considération dans toutes les analyses. La plus grande partie des logiciels statistiques permettent l'utilisation de poids, mais la définition qu'ils en font peut être quelque peu différente. Ainsi, bien que la plupart des estimations produites à l'aide des poids soient habituellement correctes, les estimations de la variabilité d'échantillonnage pourront être incorrectes. Pour calculer les variabilités d'échantillonnage exactes, on doit ainsi connaître tous les détails du plan d'échantillonnage, détails qui ne peuvent être dévoilés ici pour des fins de complexité et de confidentialité.

On propose toutefois une alternative au calcul de la variabilité d'échantillonnage de variables quantitatives, soit la méthode des groupes aléatoires. Quoique demandant un travail supplémentaire, elle devrait offrir une bonne estimation de l'erreur type désirée. Les étapes suivantes doivent être suivies:

- (i) Répartir les unités (ménages) du fichier de microdonnées en k sous-groupes de taille à peu près égale ($k = 4$ serait suffisant) et ce, à l'intérieur de chaque strate géographique (tableau 1). Les ménages étant déjà placés dans un ordre aléatoire, il suffit d'effectuer la répartition systématiquement en plaçant le premier élément et chaque k ième élément qui suit dans le premier sous-groupe et ainsi de suite.
- (ii) Calculer la valeur de la statistique désirée pour chaque sous-groupe. On doit ici multiplier le poids par la valeur k , puisque chaque sous-groupe doit représenter la population entière. Soit " z_i " la valeur de la statistique pour le sous-groupe i ($i = 1, 2, \dots, k$).

- (iii) L'erreur type de la statistique est évaluée à l'aide de la formule suivante:

$$s = \frac{1}{k(k-1)} \sum_{i=1}^k (z_i - \bar{z})^2 \quad \text{où} \quad \bar{z} = \sum_{i=1}^k \frac{z_i}{k}$$

3. Variabilité non due à l'échantillonnage

L'erreur d'échantillonnage n'est qu'une des composantes de l'erreur totale d'un sondage. Des erreurs non dues à l'échantillonnage peuvent également contribuer à cette erreur totale. Des erreurs de ce type sont introduites, par exemple, lorsque le répondant donne une mauvaise information ou qu'il ne répond pas à certaines questions (erreur de réponse), lorsqu'une unité de la population visée est oubliée ou comptée plus d'une fois (erreur d'observation) ou lors du dépouillement des données comme, par exemple, les erreurs de codage et de saisie. De plus, étant donnés des critères de confidentialité, certaines valeurs doivent être supprimées du fichier. Les mesures de la variabilité d'échantillonnage discutées aux sections précédentes ne tiennent compte que de la variabilité observée par rapport aux données du recensement. Ainsi, elles ne reflètent pas les inexactitudes introduites dans les données du recensement et de l'échantillon par les erreurs non dues à l'échantillonnage et les suppressions.

Si une faible proportion de la population est touchée par une estimation particulière, l'erreur d'échantillonnage sera la principale composante de l'erreur totale. Toutefois, plus l'estimation se rapproche de l'effectif total de la population, plus l'erreur d'échantillonnage devient négligeable, ce qui n'est pas nécessairement le cas des erreurs non dues à l'échantillonnage. En fait, plus l'estimation tend vers le total de la population correspondante, plus les erreurs non dues à l'échantillonnage deviennent importantes relativement aux erreurs d'échantillonnage.

AUTRES FACTEURS QUI AFFECTENT LA FIABILITÉ DES DONNÉES

AJUSTEMENTS AUX UNITÉS GÉOSTATISTIQUES

Les utilisateurs doivent être conscients du fait que les unités géostatistiques du recensement peuvent varier d'un recensement à l'autre. Par conséquent, celui qui utilise des données provenant de deux ou plusieurs recensements doit être au courant et tenir compte du fait que les limites géographiques des secteurs qu'il compare peuvent avoir été modifiées. Les utilisateurs qui désirent obtenir d'autres renseignements à ce sujet doivent se reporter au chapitre 3.

CHIFFRES DE POPULATION BASÉS SUR LA RÉSIDENCE HABITUELLE

Les chiffres de population présentés ici pour les différentes régions géographiques indiquent le nombre de Canadiens qui y ont leur lieu habituel de résidence, sans égard au lieu où ils se trouvaient le jour du recensement. Sont également comptés les Canadiens séjournant dans un logement de la région en cause le jour du recensement et n'ayant pas de lieu habituel de résidence ailleurs au Canada. Dans la plupart des régions, il n'y a guère de différence entre le nombre de résidents habituels et le nombre de personnes séjournant dans la région le jour du recensement. Toutefois, dans certains endroits, comme les lieux de villégiature et les régions où sont situés des camps de chantier importants, la différence entre ces deux chiffres peut être importante.

IMMIGRANTS ET PERSONNES NÉES EN DEHORS DU CANADA

Toutes les personnes nées en dehors du Canada ne sont pas nécessairement des immigrants au Canada. En effet, les personnes qui ont indiqué être nées en dehors du Canada mais qui sont maintenant des citoyens canadiens de naissance ne sont pas considérées comme des immigrants au Canada. Par conséquent, elles n'ont pas à déclarer l'année d'immigration ou leur âge à l'immigration lorsqu'elles s'établissent au Canada de façon permanente. Ces personnes ont été classées dans la catégorie des non-immigrants, comme au

recensement de 1981. Par contre, au recensement de 1971, toutes les personnes nées en dehors du Canada ont été classées comme immigrants et ont dû répondre à la question sur la période d'immigration.

MOBILITÉ

Les limites des régions géographiques sont celles au 1^{er} janvier 1986, qui est la date de référence géographique du recensement du Canada de 1986.

Les chiffres pour le total des "migrants" (un migrant est une personne qui, cinq ans plus tôt, n'avait pas son lieu habituel de résidence à l'intérieur de la subdivision de recensement (SDR) où il a été dénombré) sont additifs au niveau géographique, c'est-à-dire que le nombre de migrants au niveau du Canada est la somme du nombre de migrants au niveau provincial.

Au niveau de la SDR, il est recommandé que les utilisateurs fassent preuve de prudence lorsqu'ils utilisent des données du nombre de migrants, en particulier dans le cas des municipalités de banlieue à l'intérieur des grandes régions métropolitaines. Les chiffres du total des migrants, y compris les immigrants et les émigrants internes, peuvent être déformés en raison de la possibilité de mauvaises réponses dans des cas tels que ceux-ci: a) des répondants dans les régions métropolitaines ont déclaré la ville principale plutôt que la municipalité où ils vivaient cinq ans plus tôt (Toronto, au lieu de Scarborough, par exemple); b) les répondants ne déclarent pas un déménagement d'une autre SDR s'ils estiment qu'ils sont toujours dans la même ville principale (ils ont fait un déménagement de Toronto à Scarborough, mais ont indiqué qu'ils habitaient toujours la même municipalité); et c) les répondants ont déclaré des déménagements selon des limites désuètes.

Le concept de "migrant" est défini au niveau de la SDR. Dans le cas des régions géographiques inférieures à la SDR telles que les secteurs de dénombrement (SD) et les secteurs de recensement (SR), le lecteur est prié de noter que la distinction entre la population des migrants et des non-migrants se rattache à la SDR correspondante du SD ou du SR. Ainsi, les migrants de la SR sont les personnes qui ont déménagé d'une autre SDR, tandis que les non-migrants sont ceux qui ont déménagé à l'intérieur de la même SDR, c'est-à-dire qu'ils sont passés d'un SR à un autre ou ont déménagé à l'intérieur du même SR.

Activité, recensement du Canada de 1981

Canada	Recensement de 1981 (chiffres publiés en 1981)	Recensement de 1981 (selon le dépouillement de 1986)	Variation en %
Population active de 15 ans et plus	12,054,150	12,081,280	0.23
Personnes occupées	11,167,915	11,167,915	aucune variation
Chômeurs	886,235	913,365	3.06
Inactifs	6,555,135	6,528,005	-0.41
Population active de 15-19 ans	1,073,945	1,098,390	2.28
Personnes occupées	906,705	906,705	aucune variation
Chômeurs	167,240	191,680	14.61
Inactifs	1,229,630	1,205,190	-1.99
Population active de 20 ans et plus	10,980,205	10,982,890	0.02
Personnes occupées	10,261,210	10,261,210	aucune variation
Chômeurs	718,995	721,685	0.37
Inactifs	5,325,505	5,322,815	-0.05

Les noms et les limites de certaines subdivisions de recensement peuvent subir des modifications minimales, mais parfois importantes, au cours de la période de cinq ans entre deux recensements, et les comparaisons des données pour une région infraprovinciale donnée entre deux recensements ne seront donc pas valides si ces changements éventuels ne sont pas pris en compte.

On peut trouver des renseignements sur les changements de limites entre recensements dans la **Classification géographique type** (n° 12-573 au catalogue).

Les limites des RMR et des AR et leurs SDR composantes vont souvent subir des modifications entre deux recensements, et la comparaison des données pour des régions déterminées à ce moment-là ne sera pas valide à moins que ces changements ne soient pris en compte. Une publication contient les comparaisons des RMR et des AR de 1986 et de leurs homologues de 1981. La publication **Régions métropolitaines de recensement et agglomérations de recensement: Comparaison entre 1986 et 1981** (n° 99-105F ou E au catalogue) donne la liste des subdivisions de recensement qui composent la version 1986 de chaque RMR et AR et présente les limites correspondantes pour 1981.

NOMBRE DE SEMAINES TRAVAILLÉES

Les données concernant le nombre de semaines travaillées en 1985 doivent être utilisées avec circonspection pour les catégories 40-48 semaines et 49-52 semaines, car certains répondants ont

tendance à ne pas inclure leurs congés payés de vacances ou pour autres raisons dans leurs semaines de travail, tel que requis. Ainsi, la catégorie 49 à 52 semaines peut être sous-estimée.

ACTIVITÉ

Les concepts de l'activité au recensement n'ont pas changé entre 1981 et 1986. Par contre, le traitement des données a été modifié, ce qui a entraîné quelques différences. Au recensement de 1986, contrairement aux recensements antérieurs, on n'a posé aucune question sur la fréquentation scolaire. Cette question était utilisée pour le contrôle de la variable de l'activité, plus spécifiquement le chômage. Par conséquent, ces différences de traitement affectent la population en chômage et sont en grande partie concentrées dans le groupe d'âge des 15 à 19 ans. Le tableau ci-dessus indique l'ampleur de cet effet sur les données, au niveau du Canada.

COMPARABILITÉ ET QUALITÉ DES DONNÉES LINGUISTIQUES

Comparaison entre 1981 et 1986

Langue maternelle et langue parlée à la maison. Les mêmes questions linguistiques ont été posées lors des deux derniers recensements mais pour celles sur la langue maternelle et la langue parlée à la maison, les instructions données aux répondants ont été modifiées. En 1981, les répondants étaient invités à n'indiquer qu'une langue maternelle et qu'une langue parlée à la

maison. Malgré cela, 597,980 personnes (2.5% de la population) avaient déclaré plus d'une langue maternelle et 535,735 personnes (2.2% de la population) avaient déclaré plus d'une langue parlée à la maison.

Afin de mieux rendre compte de la réalité linguistique au Canada, cette instruction a été supprimée du recensement de 1986. En vertu des nouvelles instructions données aux répondants, une personne pouvait indiquer deux langues maternelles si elle les avait apprises simultanément et les avait parlées aussi souvent l'une que l'autre dans son enfance. De même, une personne pouvait indiquer plus d'une langue parlée à la maison si elle les utilisait actuellement aussi souvent l'une que l'autre à la maison.

Le nombre de réponses multiples s'est accru de manière importante lors du recensement de 1986 par rapport à celui de 1981. En 1986, 954,940 personnes (3.8% de la population) ont fourni une réponse multiple à la question sur la langue maternelle et 1,159,675 personnes (4.6% de la population) ont indiqué parler plus d'une langue à la maison.

Cette augmentation résulte soit des modifications apportées au questionnaire, soit des changements dans la manière dont la population répond aux questions linguistiques, soit d'un accroissement du nombre de personnes ayant appris plus d'une langue maternelle ou du nombre de personnes parlant plus d'une langue à la maison. L'augmentation des réponses multiples peut aussi résulter d'une combinaison de ces facteurs.

Lors du traitement des données de 1981, on n'a retenu et publié qu'une seule langue, même dans les cas où plusieurs avaient été déclarées. En 1986, on a accepté les réponses comprenant plus d'une langue.

Pour faciliter l'établissement des tendances entre les deux recensements, les données du recensement de 1986 ont été ajustées. Dans les cas où plus d'une langue a été indiquée, les réponses multiples ont été réparties entre les langues composantes dans les mêmes proportions qu'en 1981. Les résultats ont été publiés dans un document spécial intitulé "**Données linguistiques ajustées**", avril (1988). De plus, les données du recensement de 1981 ont été ajustées pour faire état des réponses multiples alors obtenues. Elles sont présentées au tableau 4 des publications 93-102 (langue maternelle) et 93-103 (langue parlée à la maison). Ces ajustements aux chiffres de la langue maternelle et de la langue parlée à la maison

facilitent la mise en relation des données de 1986 avec celles de 1981, mais ne rendent pas parfaitement comparables les résultats des deux recensements. Par conséquent, il faut être très prudents dans l'interprétation des changements observés entre 1981 et 1986.

Des modifications ont été apportées à la classification des langues de 1986 par rapport à celles de 1981, notamment en ce qui concerne les langues autochtones. L'annexe B du **Dictionnaire du recensement de 1986** (n° 99-101F au catalogue) nous donne une description des changements.

Langue officielle. Certains répondants ont déclaré parler l'anglais ou le français ou les deux à la maison alors qu'ils ont indiqué, à la question sur la langue officielle, ne pas pouvoir soutenir une conversation dans ces deux langues.

Dans de tels cas, au recensement de 1981, la réponse à la question sur la langue officielle était considérée comme étant erronée. Par conséquent, au cours du traitement des données, cette réponse était modifiée afin de montrer que la personne pouvait parler la (les) langue(s) officielle(s) qu'elle avait déclarée(s) à la question sur la langue parlée à la maison.

Au recensement de 1986, ces réponses n'ont pas toujours été considérées comme étant erronées. Si le répondant avait indiqué pouvoir parler une seule des langues officielles, soit l'anglais ou le français, et que cette langue correspondait à la langue maternelle de cette personne, aucune correction n'a été apportée lors du traitement. Par conséquent, ces enchaînements de réponses apparaissent tels quels dans les totalisations de 1986.

Pour de plus amples renseignements au sujet des données linguistiques, veuillez communiquer avec la Division des statistiques sociales, du logement et des familles, Statistique Canada, Ottawa, Canada K1A 0T6.

COMPARABILITÉ DES DONNÉES SUR L'ORIGINE ETHNIQUE

Comparaison des chiffres de 1981 et de 1986. Les données sur l'origine ethnique des recensements de 1981 et de 1986 ne sont pas directement comparables.

La question de 1981 portant sur l'origine ethnique qui se lisait "À quel groupe ethnique ou culturel apparteniez-vous, vous ou vos ancêtres, à votre première arrivée sur ce continent?" a été modifiée pour le recensement de 1986. Dans la version de 1986, on a retranché le segment suivant: "à votre première arrivée sur ce continent", car on a jugé qu'il n'était pas valable dans le cas des personnes d'origine autochtone. En 1986, la question a été reformulée comme suit: "À quel(s) groupe(s) ethnique(s) ou culturel(s) appartenez-vous ou vos ancêtres appartenaient-ils?"

En 1986, les répondants devaient indiquer ou préciser leur appartenance à un ou plusieurs groupes selon le cas. En outre, l'addition de deux espaces où le répondant pouvait inscrire une réponse en toutes lettres est l'une des raisons principales de l'augmentation du nombre de réponses multiples à la question sur l'origine ethnique.

De plus, l'ordre de présentation des cases à cocher en réponse à cette question est fondé sur la déclaration selon la fréquence ou les origines ethniques simples relevées en 1981. Ceci a entraîné le changement de position des cases marquées "Chinois" et "Polonais".

Compte tenu des recommandations formulées par le Comité spécial sur la participation des minorités visibles à la société canadienne dans le document intitulé: "L'égalité ça presse!" et par la Commission sur l'égalité en matière d'emploi présidée par le juge Abella, la case marquée "Noir" a été ajoutée à la question sur l'origine ethnique en 1986.

Les cases réservées aux autochtones ont également été modifiées. On a remplacé en 1986 les catégories d'Indiens inscrits ou non inscrits, utilisées en 1981, par l'expression "Indien de l'Amérique du Nord". Il convient de noter que des personnes qui ne sont pas d'origine autochtone mais qui détiennent le statut d'Indien inscrit aux termes de la Loi sur les Indiens, à la suite d'un mariage par exemple, ont pu être incluses dans les données sur les autochtones de 1981. Il est possible que ces personnes ne se soient pas identifiées comme "Indien de l'Amérique du Nord" en 1986 et, par conséquent, ne sont pas comptées dans le nombre d'autochtones en 1986. De plus, un nombre indéterminé de personnes d'origine métisse ont pu déclarer, en 1986, une réponse multiple comprenant "Indien de l'Amérique du Nord" et une ou plusieurs autres origines ethniques ou culturelles.

Réponses unique et multiple

Lorsqu'un répondant n'indique qu'une seule origine, il s'agit d'une "**réponse unique**". Par exemple, 709,585 personnes au Canada n'ont indiqué comme origine ethnique que l'Italien.

Quand un répondant fait connaître plus d'une origine, il donne une "**réponse multiple**". Au Canada, quelque 297,325 personnes d'ascendance italienne ont en outre signalé une ou plusieurs autres origines ethniques ou culturelles. Ainsi, 31,495 personnes ont donné la réponse multiple suivante: Italien et Français.

Dans la légende pour l'origine ethnique de ce profil, les origines uniques sont indiquées comme des groupes uniques. Les origines multiples apparaissent comme un groupe: origines multiples. Les 31,495 réponses multiples "Italien et Français" sont comprises dans le compte des origines multiples (6,986,345 pour le Canada).

Pour plus de renseignements sur les données touchant l'origine ethnique, veuillez communiquer avec la Division des statistiques sociales, du logement et des familles, Statistique Canada, Ottawa (Ontario) K1A 0T6, au numéro (613) 951-2574.

SOUTIEN DU MÉNAGE

Les utilisateurs des données sur le soutien du ménage, comme celles sur le sexe du soutien ou la langue maternelle de celui-ci, doivent savoir que ces données comportent des limites qui pourraient avoir une grande incidence sur l'usage et l'analyse que l'on en fait.

La variable du soutien du ménage est une variable dérivée, c'est-à-dire qu'elle résulte de la combinaison et de la manipulation des réponses que les utilisateurs ont fournies à la question sur la "personne responsable des paiements du ménage" et à celle sur le "lien avec la Personne 1". Cette variable vise à classer les familles que compte un ménage en familles principales (c'est-à-dire les familles où le soutien est aussi membre de la famille) ou en familles secondaires (c'est-à-dire les familles où le soutien n'est pas membre de la famille). Cette variable n'a pas été conçue de manière à servir d'équivalent à l'ancienne variable de "chef de ménage" à des fins analytiques.

Pendant le traitement, la variable elle-même n'a pas été considérée comme une variable devant servir lors d'éventuelles analyses. Ainsi, si un répondant inscrivait plus d'un nom sous la question relative à la "personne responsable des paiements du ménage", seul le premier était saisi; les autres étaient éliminés. De plus, si un répondant indiquait qu'aucune personne du ménage ne faisait de paiement au titre du logement, ce ménage était laissé sans famille principale, mais la Personne 1 était arbitrairement désignée soutien du ménage. Ces décisions ont été prises parce qu'il fallait avant tout classer les familles en familles principales ou secondaires et non fournir une personne repère pour le ménage.

On prévient donc les utilisateurs de ne pas faire de déductions injustifiées en se fondant seulement sur des comparaisons directes des caractéristiques des soutiens du ménage. Ainsi, il faut être prudent lorsqu'on compare les soutiens de sexe féminin et les soutiens de sexe masculin puisqu'un nombre inconnu de ceux-ci peuvent avoir été placés en second sur la liste figurant sous la question relative à la "personne responsable des paiements du ménage" et par conséquent avoir été éliminés. De la même façon, il se peut qu'une personne de l'extérieur du ménage ait parfois été remplacée par la "Personne 1" lors de la détermination du soutien du ménage, et que les données sur le sexe du soutien du ménage soient ainsi faussées.

Une interprétation erronée des résultats peut aussi se produire lorsqu'on a recours à d'autres caractéristiques du soutien, comme sa langue maternelle ou son origine ethnique, pour classer un ménage puisque ces caractéristiques peuvent différer pour les autres membres de celui-ci. On suggère que les analyses utilisant ces variables tiennent aussi compte des caractéristiques du conjoint du soutien.

TYPE DE CONSTRUCTION

Les utilisateurs des données sur le type de construction résidentielle devraient être bien conscients que ces données présentent certaines limites. Une enquête préliminaire révèle en effet qu'elles comportent des limites qui pourraient avoir pour effet d'en diminuer la qualité, les voici:

- (1) Lors du recensement de 1986, on a enregistré un taux plus élevé de non-réponse qu'en 1981 pour la question sur le type de construction résidentielle (2.3% comparativement à 0.5%). Cette augmentation du taux de non-réponse ne devrait avoir qu'une faible incidence sur la qualité globale des données, sauf pour un

nombre limité de régions géographiques où l'on peut trouver une proportion plus grande de non-réponses. Il faut aussi signaler que ce type de renseignements était fourni par le recenseur en 1986 tandis qu'il était transmis par le répondant du ménage en 1981.

- (2) Des baisses marquées entre les recensements de 1981 et de 1986 ont été remarquées dans toutes les provinces pour les habitations mobiles et les autres types de logements mobiles. On croit que ce phénomène est attribuable au fait qu'un certain nombre d'habitations mobiles ont été classées par erreur parmi les autres types de construction, principalement les logements individuels non attenants. Pour les régions géographiques plus étendues, on ne s'attend pas à ce que cette erreur ait des conséquences importantes pour les autres catégories de logement en raison du nombre relativement faible d'habitations et de logements mobiles.
- (3) Les données sur les appartements des immeubles de moins de cinq étages présentent certaines différences par rapport au recensement de 1981, particulièrement au Québec et plus précisément à Montréal. De plus, un fort surdénombrement des duplex, des maisons jumelées et des maisons en rangée en 1981 a entraîné de très fortes baisses de ces types de logements dans certaines provinces en 1986. Une analyse historique préliminaire indique que les chiffres de 1986 étaient assez réalistes.

DONNÉES SUR LE REVENU

Le concept du revenu total au recensement de 1986 comprend, pour la première fois, les crédits d'impôt fédéral pour enfants. Le recensement de 1986, comme en 1981, n'a pas recueilli de données sur le revenu des pensionnaires d'institution. Les statistiques sur le revenu des familles et des ménages sont calculées seulement pour les personnes vivant dans des ménages privés.

Les statistiques du recensement sur le revenu sont sujettes à la variabilité d'échantillonnage. Bien qu'une telle variabilité d'échantillonnage puisse être assez petite pour de grands groupes de population, ses effets ne peuvent être ignorés dans le cas de très petits sous-groupes de population dans une région ou dans une certaine catégorie. Il en est ainsi parce que, toutes choses étant égales par ailleurs, plus grande est la taille de l'échantillon, plus petite est l'erreur. C'est pour cette raison que les données

publiées sur le revenu dans les régions sous le niveau provincial où la population non institutionnelle était de moins de 250 habitants ont été supprimées. Par contre, là où les statistiques ne sont pas supprimées mais sont néanmoins fondées sur des totaux relativement petits, nous conseillons fortement aux lecteurs d'être prudents dans l'utilisation et l'interprétation de ces statistiques.

Catégorie de revenu

Le catégorie de revenu se rapporte à la situation des familles économiques et des personnes seules selon les seuils de faible revenu établis par Statistique Canada. Ces seuils sont déterminés séparément pour les familles de taille différente et celles vivant dans des secteurs ayant des niveaux d'urbanisation différentes. Pour le recensement de 1986, ils sont fondés sur les seuils de faible revenu révisés (1978) qui ont d'abord été estimés à la lumière de l'enquête nationale de 1978 sur les dépenses des familles, puis mis à jour en 1985 en fonction de la variation de l'indice des prix à la consommation depuis 1978. La matrice des seuils de faible revenu de 1985 est indiquée au tableau 1 ci-dessous.

Aux fins des statistiques sur le faible revenu, les familles économiques et les personnes seules résidant au Yukon, dans les Territoires du Nord-

Ouest et dans les réserves indiennes ne sont pas prises en compte. Les seuils de faible revenu s'inspiraient de certaines tendances "dépenses-revenu" que les données d'enquête ne fournissaient pas pour la population entière.

Pour déduire la fréquence des unités à faible revenu, on utilise le critère "taille de la région", lequel diffère selon qu'il s'agit du recensement ou de l'enquête sur les finances des consommateurs. Pour le recensement, la taille de la région est en effet déterminée par la population totale des secteurs de dénombrement urbains contigus (un secteur de dénombrement est considéré comme urbain en fonction de la densité de la population). Pour l'enquête sur les finances des consommateurs, on prend des RMR ou des AR et on détermine la taille de la région selon la population totale comprise dans les limites de ces RMR ou AR. Les répercussions globales de cette différence sont négligeables.

Pour plus de détails concernant les aspects conceptuels et le champ d'observation, voir la publication Revenu de la famille, familles économiques, n° 93-918 au catalogue. (Voir également la publication Répartition du revenu selon la taille du Canada, 1985, N° 13-207 au catalogue.)

Tableau 1: Seuils de faible revenu des unités familiales, 1985

Taille de l'unité familiale	Taille du secteur de résidence				
	500,000 et plus	100,000 - 499,999	30,000 - 99,999	Petites régions urbaines	Régions rurales
dollars de 1985					
1 personne	10,233	9,719	9,117	8,429	7,568
2 personnes	13,501	12,815	11,956	11,093	9,891
3 personnes	18,061	17,115	15,996	14,880	13,244
4 personnes	20,812	19,779	18,490	17,200	15,310
5 personnes	24,252	22,963	21,415	19,952	17,803
6 personnes	26,488	25,026	23,393	21,758	19,436
7 personnes ou plus	29,155	27,606	25,801	23,994	21,415

Arrondissement et ajustement des revenus élevés et des pertes

Au cours de la planification du présent fichier de microdonnées, il a été jugé essentiel que les procédures utilisées rendent impossible le lien entre un particulier, une famille ou un ménage et son revenu. Le revenu des particuliers dans les ménages sélectionnés pour faire partie du fichier de microdonnées a donc été soumis à la procédure d'arrondissement et d'ajustement des revenus décrite ci-dessous.

Le revenu de chaque membre des ménages compris dans le fichier a fait l'objet de deux opérations distinctes. On a d'abord arrondi, aux limites spécifiées au tableau 2, les salaires et traitements, le revenu provenant d'un travail autonome (agricole et non agricole), le revenu de placements, les pensions de retraite, les autres revenus en espèces et le revenu total. Cet arrondissement a fait que la somme des revenus des diverses sources ne correspondait plus nécessairement au revenu total. Ces incohérences ont été corrigées selon la procédure d'ajustement décrite au tableau 3. Une fois les données des enregistrements sur les particuliers arrondies et ajustées, les variables du revenu des familles et des ménages du fichier de

microdonnées, c'est-à-dire le revenu d'emploi, les transferts gouvernementaux, le revenu de placements, tous les autres revenus en espèces et le revenu total, ont été calculées.

Les tableaux qui suivent indiquent le nombre d'enregistrements affectés par cette procédure et schématisent l'incidence de celle-ci sur le revenu des ménages.

Le tableau 4 donne une répartition des particuliers dont les revenus d'une ou de plusieurs sources et (ou) le revenu total se situent hors des limites imposées par la confidentialité.

Le tableau 5 présente sommairement au niveau de l'échantillon les variations, attribuables à la procédure d'arrondissement/ajustement, du revenu agrégé et du revenu moyen du ménage selon la source.

Le tableau 6 donne une répartition du revenu agrégé pondéré du ménage en 1985 selon la province, à partir de la base de données principale du recensement et du Fichier de microdonnées à grande diffusion.

Tableau 2: Limites de revenu inférieure et supérieure

(1) Pour tous les particuliers de 15 ans et plus dans l'échantillon du fichier de microdonnées sur les ménages, les revenus des sources suivantes ont été arrondis aux limites inférieure et supérieure:			
a) Salaires et traitements			
b) Revenu provenant d'un travail autonome			
c) Revenu de placements			
d) Pensions de retraite			
e) Autres revenus en espèces			
(2) Les limites s'établissaient comme suit:			
		Limite <u>inférieure</u>	Limite <u>supérieure</u>
a)	Femmes de toutes les régions et hommes de la région de l'Atlantique	-\$30,000	\$100,000
b)	Hommes de toutes les autres régions	-\$50,000	\$140,000
(3) Les montants supérieurs aux limites indiquées au point (2) ci-dessus ont été arrondis aux limites.			
(4) Dans les cas où le revenu total était supérieur à la limite, la première étape a été d'arrondir celui-ci à la limite applicable.			
(5) Pour que la somme des sources corresponde au revenu total, les enregistrements sur les particuliers ont été soumis à la procédure d'ajustement décrite au tableau 3.			

Tableau 3: Ajustements apportés afin d'éliminer les incohérences produites par l'arrondissement

Après l'arrondissement des revenus des différentes sources applicables et (ou) du revenu total tel que décrit au tableau 2, les revenus de chaque source et le revenu total ont été soumis aux ajustements décrits ci-dessous pour que la somme des revenus des différentes sources corresponde au revenu total:

- I Ajustement des revenus des différentes sources
- (1) Si $A > 0$ et $B > 0$ et $A < B$ alors $S_f = (S_i) (C/D)$

(2) Si $A < 0$ et $B < 0$ et $A > B$ alors $S_{ef} = S_e + A - B$

(3) Aucun ajustement dans tous les autres cas.
- II Ajustement du revenu total
- (1) $Y =$ Somme des revenus des différentes sources (après les ajustements apportés au point I ci-dessus et y compris les transferts)

(2) $Y = 1$ si la somme des revenus ajustés des différentes sources et des transferts $= 0$
- A = Revenu total après arrondissement
- B = Somme des revenus des différentes sources après arrondissement
- C = A moins les transferts
- D = B moins les transferts
- S_i = Salaires, revenu provenant d'un travail autonome, revenu de placements, pensions de retraite et autres revenus en espèces après arrondissement
- S_f = Salaires, revenu provenant d'un travail autonome, revenu de placements, pensions de retraite et autres revenus en espèces finals au FMGD
- S_e = Revenu provenant d'un travail autonome après arrondissement
- S_{ef} = Revenu final provenant d'un travail autonome au FMGD
- Y = Revenu final total au FMGD

Les revenus du ménage et de la famille ont été calculés en additionnant les revenus des particuliers dans le ménage et la famille après les ajustements décrits ci-dessus.

Tableau 4: Répartition (échantillon non pondéré) des particuliers dont le revenu, en 1985, dépassait les limites (1) inférieure ou supérieure, fichier de microdonnées à grande diffusion sur les ménages et les logements, recensement de 1986

Revenus hors limites	Chiffre de l'échantillon	%
Une source	436	74.7
Salaires et traitements	182	31.2
Revenu provenant d'un travail autonome	181	31.0
Revenu de placements	69	11.8
Pensions de retraite	2	.3
Autres revenus en espèces	2	.3
Deux sources	7	1.2
Salaires et revenu de placements	5	.9
Revenu provenant d'un travail autonome et revenu de placements	2	.3
Revenu total seulement	140	24.0
Total (2)	584	100.0

(1) Voir le tableau 2 pour les limites.
(2) Dans les 562 ménages, 584 personnes avaient des revenus hors des limites spécifiées.

Tableau 5: Nombre de ménages, leur revenu moyen et agrégé original et modifié en 1985 selon la source, Fichier de microdonnées à grande diffusion sur les ménages et les logements, recensement de 1986

Nombre, revenu agrégé et revenu moyen	Revenu d'emploi	Revenu de placements	Pensions de retraite et autres revenus en espèces	Transferts gouverne- mentaux	Revenu total
1. Nombre d'enregistrements	115,347	115,347	115,347	115,347	115,347
2. Revenu agrégé (\$'000)					
a. Original, total	3,111,602	251,486	146,169	435,601	3,944,859
b. Modifié (562 ménages)	-27,662	-9,456	-2,204	0	-39,323
c. Final, total	3,083,941	242,030	143,965	435,601	3,905,537
d. Taux de variation - (b/a)	-0.9	-3.8	-1.5	0.0	-1.0
3. Revenu moyen par ménage					
a. Original, total	26,976	2,180	1,267	3,776	34,200
b. Modifié (562 enregistrements)	-49,220	-16,826	-3,922	0	-70,428
c. Modification globale	-240	-82	-19	0	-343
d. Final, total	26,736	2,098	1,248	3,776	33,859
e. Taux de variation (d/a)	-0.9	-3.8	-1.5	0.0	-1.0

Tableau 6: Répartition (montant et pourcentage) du revenu agrégé des ménages en 1985 selon la province, recensement de 1986 et estimations provenant du Fichier de microdonnées à grande diffusion sur les ménages et les logements du recensement de 1986

Province	Dollars		%		Différence FMGD/ Recensement
	Recensement ⁽¹⁾	FMGD Recensement ⁽¹⁾	FMGD	FMGD	
	(\$'000,000)			%	
Terre-Neuve	4,610.4	4,567.2	1.5	1.5	-0.9
Île-du-Prince-Édouard	1,155.5	1,143.9	0.4	0.4	-1.0
Nouvelle-Écosse	9,068.3	8,857.1	2.9	2.9	-2.4
Nouveau-Brunswick	6,657.6	6,672.3	2.2	2.2	0.2
Québec	73,067.7	72,021.8	23.7	23.8	-1.5
Ontario	122,496.3	120,480.2	39.8	39.8	-1.7
Manitoba	11,952.8	11,560.0	3.9	3.8	-3.4
Saskatchewan	11,029.6	10,893.9	3.6	3.6	-1.2
Alberta	30,766.5	30,214.8	10.0	10.0	-1.8
Colombie-Britannique	36,414.8	35,703.4	11.8	11.8	-2.0
Yukon/Territoires du Nord-Ouest	845.3	833.6	0.3	0.3	-1.4
Canada	308,064.8	302,948.7	100.0	100.0	-1.7

(1) Base de données du recensement de 1986 sans arrondissement aléatoire.