

MICRODATA FILE - 1995 INCOME

KEY FILE

SAMPLE DESIGN AND IMPLICATIONS

The sample employed for the **Survey of Consumer Finances and Household Facilities and Equipment Survey** is the Labour Force Survey sample. This multi-stage stratified clustered probability sample is described in detail in **Methodology of the Canadian Labour Force Survey, 1984-1990**, Catalogue No. 71-526 and the feature articles in **The Labour Force (October 1994, December 1994 and January 1995)** Catalogue No. 71-001. For a more general description, see **Guide to Labour Force Survey Data**, Catalogue No. 71-528.

Briefly, this sample is designed to represent approximately 98% of the population. Excluded population groups are:

1. Residents of the Yukon and Northwest Territories;
2. Residents of Indian reserves;
3. Residents of military barracks;
4. Inmates of institutions such as prisons, penitentiaries, jails, reformatories, mental hospitals, tuberculosis hospitals, sanatoria, orphanages and homes for the aged.

Because of varying sampling and response ratios, each record on the file has a weight attached to it. This weighting factor reflects the sample design and incorporates the inverse of the sampling ratio (which varies significantly by geographic area) and differential response rates for households, among other things. For further information on the weighting procedure, see Catalogue No. 13-207, as well as Catalogues No. 71-526, 71-528 and the issues of Catalogue No. 71-001 mentioned above.

FICHIER DE MICRODONNÉES - REVENU DE 1995

FICHIER CLÉ

PLAN DE SONDAGE ET IMPLICATIONS

L'échantillon utilisé pour l'**Enquête sur les finances des consommateurs** est celui de l'Enquête sur la population active. Cet échantillon aléatoire à grappes stratifiées à plusieurs degrés, est décrit en détail dans la publication **Méthodologie de l'enquête sur la population active du Canada, 1984-1990** (n° 71-526 au catalogue) et dans l'article du mois paru dans **La population active (octobre 1994, décembre 1994 et janvier 1995)** n° 71-001 au catalogue. Pour une description plus générale, voir le **Guide d'utilisation des données de l'enquête sur la population active**, n° 71-528 au catalogue.

En somme, cet échantillon est conçu de manière à représenter environ 98% de la population. Sont exclus les groupes suivants:

1. Les résidents du Yukon et des Territoires du Nord-Ouest;
2. Les résidents des réserves indiennes;
3. Les résidents de baraquements militaires;
4. Les pensionnaires d'institutions comme les prisons, les pénitenciers, les maisons de correction, les hôpitaux pour malades mentaux, les hôpitaux pour tuberculeux, les sanatoriums, les orphelinats et les foyers pour personnes âgées.

En raison de la variation des fractions de sondage et de taux de réponse, nous avons attribué un poids à chaque dossier du fichier. Ce coefficient de pondération tient compte du plan de sondage et comporte l'inverse de la fraction de sondage (qui varie beaucoup d'une région géographique à l'autre) et, entre autres, différents taux de réponse pour les ménages. Pour obtenir de plus amples renseignements sur la méthode de pondération, voir la publication n° 13-207 au catalogue ainsi que les publications n°s 71-526, 71-528 et les parutions de n° 71-001 au catalogue cités plus haut.

KEY FILE**FICHIER CLÉ****SAMPLE DESIGN AND IMPLICATIONS -
continued**

The sample design and weighting procedure have important implications for data use. Each variable on the file must be multiplied by the weight in order to produce valid statistical estimates of population parameters such as means, medians, modes, estimated numbers and aggregates. However, the appropriateness of using these sample survey weights in regression analysis is another issue. A useful summary of this issue is presented in DuMouchel and Duncan (1983) "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples". Journal of The American Statistical Association, Volume 78, December 1983, pp 535-543.

Since a multi-stage stratified clustered sample design was employed, standard error estimates based on the usual simple random formula generally underestimate the true standard errors. For an illustration of the differences, see the uncatalogued paper, "**An Investigation of Variance Estimation in Family Expenditure Surveys**" and the attached bibliography. See also the standard errors of average incomes (which estimate the sampling error) published in Survey of Consumer Finance, Catalogues No. 13-207 and 13-208. Standard errors of percentage distributions for individuals and family units by size are published in Catalogue No. 13-207.

The information necessary to estimate the standard errors is not given on this file. Most regression programs are not capable of considering design effects, even if the relevant information were available. However, generalized programs such as SAS, BMDP and SPSS produce coefficients using weighted least squares. The standard errors calculated by SPSS are much too low because the sum of the weights is used for the degrees of freedom. This can be corrected by dividing the weight on each record by the sum of the

PLAN DE SONDAGE ET IMPLICATIONS - suite

Le plan de sondage et la méthode de pondération comportent d'importantes implications à l'égard de l'utilisation des données. Il faut multiplier chaque variable du fichier par le poids pour obtenir des estimations statistiques valides des paramètres démographiques, comme des moyennes, des médianes, des modes et des estimations d'agrégats. Quoi qu'il en soit, l'intérêt que présente l'utilisation de ces poids de sondage dans l'analyse de régression est un autre sujet. L'article intitulé "Using Sample Survey Weights in Multiple Regression Analysis of Stratified Samples", de DuMouchel et Duncan (1983). Journal of the American Statistical Association, Volume 78, December 1983, pp 535-543, présente un sommaire utile sur la question.

Comme nous avons utilisé un échantillon à grappes stratifiées à plusieurs degrés, les estimations de l'erreur type calculées selon la formule habituelle d'échantillonnage aléatoire simple sous-représentent généralement les erreurs types vraies. Pour une illustration des différences voir le document hors catalogue intitulé "**Estimation de la variance dans les enquêtes sur les dépenses des familles**". Voir également les erreurs types des revenus moyens (qui estiment l'erreur d'échantillonnage) publiées dans les bulletins n° 13-207 et 13-208 de l'Enquête sur les finances des consommateurs. Les erreurs types des répartitions en pourcentage pour les personnes et les unités familiales selon la taille figurent dans la publication n° 13-207 au catalogue.

Les renseignements dont nous avons besoin pour estimer les erreurs types ne figurent pas dans ce fichier. La plupart des programmes de régression ne peuvent pas prendre en considération les effets du plan. Même si les renseignements pertinents sont connus. Toutefois, les programmes généraux comme le SAS, le BMDP et le PSSS produisent des coefficients au moyen des moindres carrés pondérés. Les erreurs types calculées d'après le PSSS sont beaucoup trop faibles parce qu'on y utilise la somme des poids pour

KEY FILE**SAMPLE DESIGN AND IMPLICATIONS -
concluded**

weights and multiplying by the sample size. This transformation will produce standard errors equivalent to those produced by SAS and BMDP. Specific requests to produce standard errors of estimates can be handled on a cost-recoverable basis. If this procedure is not followed, we suggest that users of regression packages be generally conservative in interpreting measures of statistical significance from these packages - perhaps using a factor of 2 or 3 to inflate the standard errors produced, since the standard errors in these packages are generally of the simple random type.

For further information on sampling and non-sampling errors associated with the survey, see Catalogue No. 13-207.

FICHIER CLÉ**PLAN DE SONDAGE ET IMPLICATIONS - fin**

établir les degrés de liberté. On peut remédier à la situation en divisant le poids de chaque dossier par la somme des poids et en multipliant par la taille de l'échantillon. On obtiendra ainsi des erreurs types équivalentes à celles que produisent le SAS et le BMDP. Nous pourrions répondre à ceux qui nous demanderont de produire des erreurs types particulières et ce, moyennant le recouvrement des frais. S'ils ne se prévalent pas de cette possibilité, nous suggérons aux utilisateurs qui se servent de programmes généraux de faire preuve de conservatisme lorsqu'il s'agit d'interpréter les mesures de la signification statistique à partir de ces programmes - et d'appliquer peut-être un facteur de 2 ou 3 pour gonfler les erreurs types produites - puisque les erreurs types dans ces programmes sont généralement de caractère aléatoire simple.

Pour de plus amples renseignements sur les erreurs d'échantillonnage et les erreurs non dues à l'échantillonnage dans l'enquête, voir la publication n° 13-207 au catalogue.