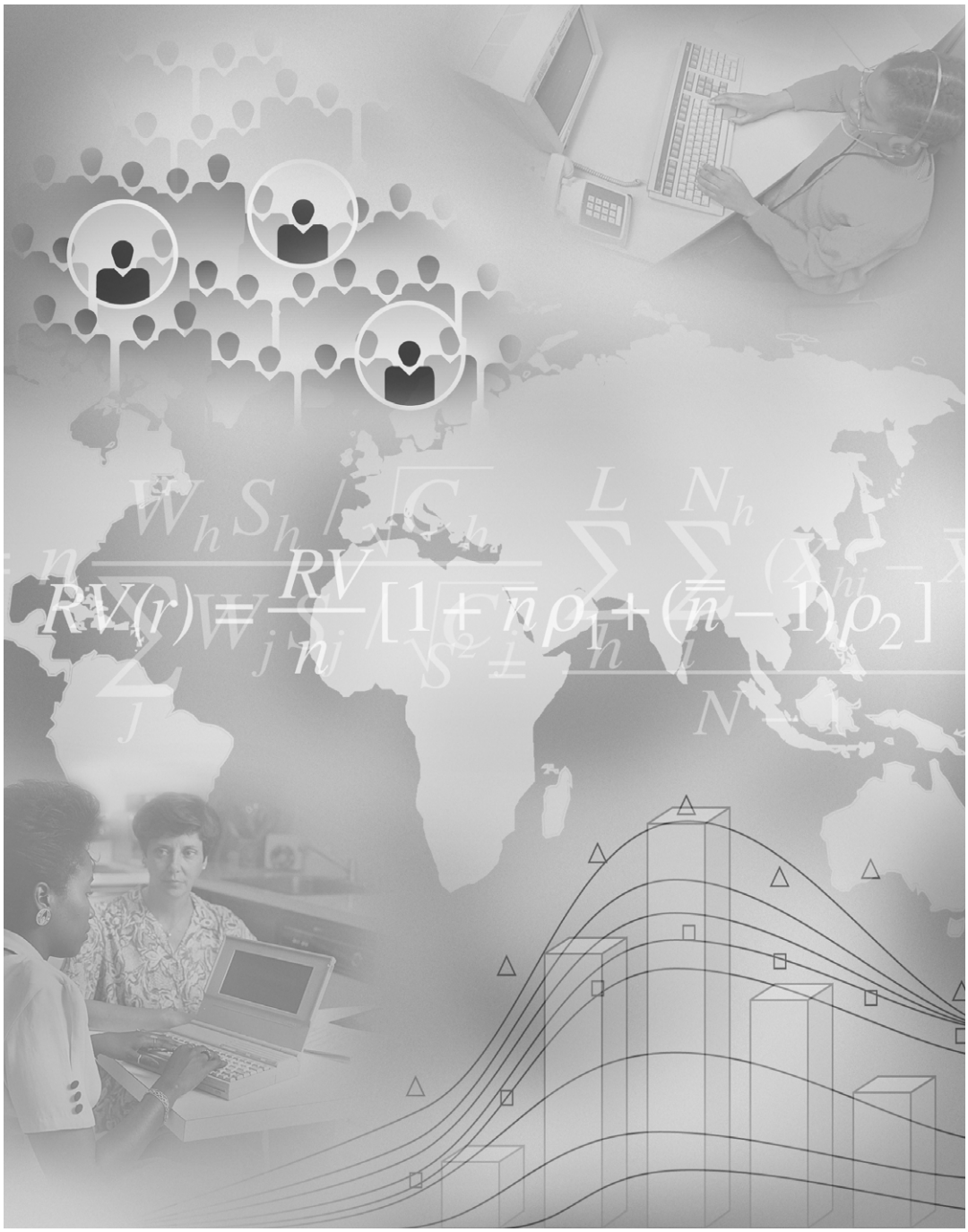
the Survey Statistician

The Newsletter of the International Association of Survey Statisticians

No. 43 December 2000



INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

International Statistical Institute

ASSOCIATION INTERNATIONALE DES STATISTICIENS D'ENQUETES

Institut International de Statistique

Editors
Leyla Mohadjer
Jairo Arrow

Section Editors
Gordon Brackstone — Country Reports
Vijay Verma — Question/Answer*
James Lepkowski — Software Review

Production Editor
Véronique Lieber

Circulation
Claude Olivier
Anne-Marie Vespa-Leyder

The Survey Statistician is published twice a year in French and English by the International Association of Survey Statisticians and distributed to all its members. Information for membership in the Association or change of address for current members should be addressed to:

Secrétariat de l'AISE/IASS
c/o INSEE-CEFIL
Att. Mme Claude Olivier
3, rue de la Cité
33500 Libourne - FRANCE
E-mail: claude.olivier@insee.fr

Comments on the contents or suggestions for articles in *The Survey Statistician* should be sent via e-mail to Mohadjl1@Westat.com or mailed to:

Leyla Mohadjer
Westat
1650 Research Blvd., Room 482
Rockville, MD 20850 - USA

In This Issue

No. 43, December 2000

In This Issue (cont.)

No. 43, December 2000

* Due to the heavy volume of special census articles, the Question/Answer Section is moved to future editions. Please send your questions (especially new practical ones) to:

Vijay Verma
105 Park Road, Teddington
Middlesex TW11 0AW, U.K.
Tel and fax: (44 20) 8977 2601
E-mail: vjverma@CompuServe.Com

Letter from the President

As I was finishing this column, sad news reached me of the death of our colleague Professor Leslie Kish. Leslie was President of IASS from 1983–85 and for many years he served superbly and tirelessly as author of the Q&A section in this newsletter. Without question, he was one of the true giants of our field, influencing generations of survey statisticians around the world through his books and articles, his lectures, and his obvious leadership qualities. Leslie was 90 and even up until recent times his sharp mind was active and undiminished at international conferences and other forums.

Statistics lost another giant last August with the death of Professor John Tukey. John was not primarily a survey statistician, nor was he an IASS member. Yet his work on computer "software" (indeed, he coined the term!) and on the "jackknife" (another term he coined!) method for the analysis of survey data profoundly affect all survey statisticians working today and, I suspect, into the foreseeable future. During the last two decades, John turned his towering intellect to the problem of census undercount, and he contributed a number of key ideas that may help reduce undercount through sampling and statistical analysis.

Professor Seymour Sudman, another of our association's most distinguished survey methodologists, died suddenly last May. Seymour was a highly productive and practical scientist who influenced all corners of the IASS domain, including sampling, response error theory, questionnaire design, and the application of cognitive psychology to survey design. He was a tremendous colleague and collaborator, and author or editor of 19 books and over 200 articles and papers, including *Response Effects in Surveys* and *Asking Questions*.

Others will write the complete obituaries for Kish, Tukey, and Sudman, but I wanted to briefly acknowledge their passing here. I personally enjoyed working with each of them, and I think it is important for all survey statisticians to remember the leaders who helped build the field we work in and benefit from today.

One of the most important activities of your association is the development and presentation of professional conferences and other training opportunities. Since the time of my last letter, the International Conference on Establishment Surveys was very successfully held in Buffalo, New York. Also, the 4th Conference on Methodological Issues in Official Statistics was successfully held in Stockholm, Sweden. My sincere thanks to John Kovar and other members of his Buffalo organizing committee and to Anders Christianson and other members of his Stockholm organizing committee for jobs well done.

Two forthcoming conferences are now in advanced planning stages. First, the Workshop on Labour Force Surveys will be held in Libourne, France, April 3-6, 2001. IASS Vice President Farhad Merhan is organizing this conference that will focus on methods and issues in small domain estimation. Second, Simona Balbi and Vincenzo Esposito are chief organizers for an IASS-IASC summer school entitled, Knowledge Discovery & Large Marketing Surveys – Design & Analysis, to be held in Naples, Italy, June 18-30, 2001. The steering committee includes IASS representatives Oreste Assereto, Martin Collins, and Jean-Michel Gautier. My thanks to all.

Two additional conferences are in the early formative stages. The International Conference on Questionnaire Development, Evaluation, and Testing is now being planned for fall 2002, with a monograph to follow. This proposed conference is in the same line of conferences as the one recently concluded in Buffalo and is to be co-sponsored by IASS, the Survey Research Methods Section of the American

Statistical Association, and others. Jennifer Rothgeb – jennifer.m.rothgeb@ccmail.census.gov – who is chair of the organizing committee is the contact. Hans Bay is organizing a conference entitled, Quality, Effectiveness, and Comparability of Surveys, to be held in Copenhagen, Denmark, August 25-28, 2002. On the other hand, Faisal Awartani informs me the previously announced regional conference Sampling Workshop in the Arab World has been cancelled.

IASS plays an active role in these conferences, either through participation in the organizing committee, through funding support, or both. I would also like to call your attention to two other conferences that IASS members are organizing but that IASS as an organization is not formally sponsoring. First, the International Conference on Quality in Official Statistics is now being planned for May 14-15, 2001, in Stockholm, Sweden. (For details see www.q2001.scb.se.) Second, Statistics Canada's symposium Achieving Data Quality in a Statistical Agency: A Methodological Perspective will be held in Ottawa, Canada, October 17-19, 2001. (For details see www.statcan.ca/english/conferences/symposium2001/index.htm.) Clearly, the general theme of "data quality" weighs on the minds of survey statisticians around the world and features in several of the forthcoming conferences.

Please get out and support these various conferences through your active participation. IASS Scientific Secretary Dan Kasprzyk is responsible for our scientific conferences and training. If you have an idea for a conference in your region and would like help getting started, please contact Dan at daniel_kasprzyk@ed.gov.

In the June issue of this newsletter, I observed the 2000 Census in the USA and urged other countries to contribute articles about their census. I am very pleased to note that a number of our country representatives have now contributed interesting census articles, including Algeria, Canada, China, Finland, France, Hungary, the Philippines, Poland, South Africa, and Spain. Many of these articles appear in this issue, and the remainder will appear in the June 2001 issue. Thanks to each of these country representatives. But I would like to followup and rechallenge other member countries to submit census articles. Within the next couple of years, *The Survey Statistician* should publish a complete record of the world's millennial censuses.

Speaking of country representatives, I would like to welcome Jairo Arrow, the new representative for South Africa, and Geoff Lee, the new representative for Australia. IASS Vice President Farhad Merhan, who coordinates the country representative program for the association, has expressed concern to me that we need more involvement from many of the countries. I ask each of you to accept his concern as a personal responsibility. Check the IASS web site or directory of members to see who is your country representative. Do you have one? Volunteer to help your representative. If you do not have a representative, volunteer to become one by contacting Mr. Merhan. It is not always easy to maintain a cohesive professional organization on a worldwide basis, and our system of country representatives is one important means we have of providing two-way communication and services to our members.

Having just mentioned the web site, I would like you to know that Fred Vogel is its new editor. I am very glad to have him on board, because a vibrant, up-to-date web site is another very important means of communicating with members in the world today. If you have improvements or comments on the web site, please send them to Fred at fvogel@nass.usda.gov.

In the last issue of this newsletter, I mentioned the IASS jubilee volume, recently named *Landmark Papers in Survey Statistics*. The content of the volume is now complete. Fritz Scheuren and Mike Brick have taken on responsibility for producing the volume. My thanks go to them, and I am confident that we will all be delighted with a finished product in the near future.

I must say I was certainly delighted with the last issue of this newsletter, with its new cover and format. Many thanks to Leyla Mohadjer and Jairo Arrow, the editors, for bringing these improvements to us.

One of the new sections in the revamped newsletter is entitled "International Flavors." This section is intended to present lighter material, such as recipes from around the world, statistical poems and jokes, and other tasteful material that would appeal to a diverse international audience. We need your help

to make this section a success. Please be a contributor and send your regular contributions and suggestions to Leyla at mohadjl1@westat.com.

The deadline for the 2001 Cochran-Hansen Prize is upon us. Please see our web site www.cbs.nl/isi/survey.htm for information about how to apply. Despite the high honor and value of the prize—a free trip to the Seoul Session and 500 Euros for books and journal subscriptions—the Secretariat has not received many applications. Please encourage qualified young statisticians to apply.

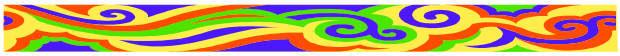
The IASS proceedings of the Helsinki session were recently disseminated to all members. My sincere thanks to Sue Linacre and her team in Australia for once again sponsoring and producing this volume for our members.

By the time you are reading these words, the Seoul Session will only be about 8 to 9 months away. I hope most of you are making plans to attend this meeting, both for its scientific and cultural aspects. Scientific Secretary Dan Kasprzyk has organized a varied and interesting program of short courses, in collaboration with local organizer, Dr. Kay-O Lee. This program is fully described in a separate article in this issue. I am especially pleased to observe that the UN Statistics Division will provide a limited number of travel and training grants for transition and developing country "young scholars." Interested members should contact Dan for the proposal requirements. IASS Program Chair David Binder has developed or co-developed 13 outstanding and timely invited paper sessions, also described in a separate column in this issue. We will organize an IASS mixer at the time of the IASS General Assembly. Please participate in as many of these activities as you can. Finally, I think it is not too late to formulate and propose contributed papers and sessions. Send in your papers now to make Seoul the biggest and best program ever.

Finally, I wish to remind members of an important, recent change in our executive office. Our new Executive Director is Christophe Lefranc, who succeeds former director Benoît Riandey, now Deputy Executive Director. For member services, Christophe can be reached at christophe.lefranc@insee.fr. I have been working with Christophe since September, and I am confident you will be pleased with the dedication and energy he is already bringing to his new job.

Best wishes,

Kirk Wolter



In Memoriam

Leslie Kish

1910-2000

Leslie Kish, a founding member of IASS, died on October 7 in Ann Arbor, Michigan. He was 90. Professor Kish was a dedicated supporter of IASS who served the Association in many ways. He was Vice President in 1977-79 and President in 1983-85. He played the major role in initiating the workshops held in conjunction with the ISI biennial meetings, and his Question/Answer Section in the *Survey Statistician* was so well-received that his contributions have been reprinted as a separate IASS publication.

Kish was born in 1910 in Poprad, which was in the Austro-Hungarian empire at that time, and he moved with his family to the United States in 1926. His father's sudden death within that year meant that he had to work during the day and complete his high school and college education at night school. In 1937, he interrupted his studies to volunteer for the International Brigade to fight for the Loyalists in the Spanish Civil War. He returned to the United States in 1939, completed his undergraduate degree in mathematics, and went to work for the Bureau of the Census and then the Division of Program Surveys at the US Department of Agriculture. When the US entered the Second World War, he volunteered for the armed forces and served as a meteorologist in the US Air Force. In 1945 he returned to the Division of Program Surveys, and then in 1947 moved with several of his colleagues to found the Institute for Social Research at the University of Michigan. He remained at ISR and on the University of Michigan faculty until his retirement in 1981, at which point he became an Emeritus Professor. After retirement he remained extremely active in the profession until his death, publishing many articles, delivering speeches, and consulting throughout the world.

Kish was a prolific writer on the theory and practice of survey sampling. In addition to a large number of papers reporting his innovative research, he authored three books. *Survey Sampling*, published in 1965 and now a Wiley Classic, is still widely read around the world for its encyclopedic treatment of the practice of survey sampling. His *Statistical Design for Research* (Wiley, 1987) provides many innovative ideas based on his extensive practical experience and his *Sampling Methods for Agricultural Surveys* (FAO, 1989) is a valuable guide to sampling practice for such surveys written for an international readership, and especially for statisticians in developing countries.

Kish devoted considerable energies to training survey statisticians throughout the world. He took great pride in the summer Sampling Program for Foreign Statisticians that he started at the University of Michigan in 1961 and that continues to thrive. In the past 40 years the program has trained about 500 samplers from 105 countries, a number of whom now hold senior positions in their government's statistical agencies. Kish visited many countries to advise on sample designs and there was nearly always a graduate of his program to meet him. When Frisbees were a novelty, he frequently took several of them on his travels to leave behind with those who enjoyed them as much as he did.

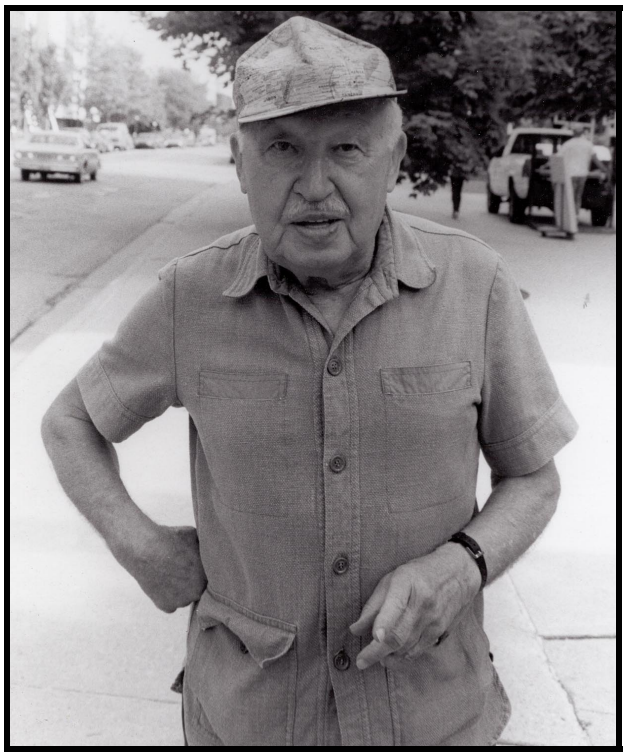
In July, at a celebration of Kish's 90th birthday in Ann Arbor, the Leslie Kish International Fellows Fund was established at the University of Michigan to provide financial support to enable students from developing countries to attend the Sampling Program. As a mark of their esteem for him and for his program, family, friends, and former students have already pledged $200,000 to the Fund.

Kish was accorded many of the highest honors from learned societies and universities, including Honorary Fellow of the ISI. He remained, however, down-to-earth, approachable by all. One of his favorite sayings was "Keep young by being curious,

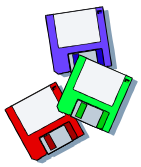
and have young friends". He followed that maxim throughout his life.

Kish was full of energy and he led a very full, productive and happy life. In addition to survey sampling, he had a great enthusiasm for many subjects, including art, music, theater, literature, history, philosophy, politics, science, the environment, and sports. He was always concerned with improving the conditions of the world's population. He much enjoyed entertaining, which he and his wife Rhea did so well. He also enjoyed dinners and conversations with his many friends in good restaurants that he had discovered. He has a loving family and is survived by Rhea, his wife of 53 years, his daughters Carla and Andrea Kish, his son-in-law Jon Stephens, his sister Magda Bondy, and his beloved granddaughter Nora Leslie Kish Stephens.

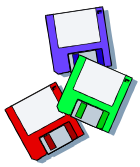
(For information about the Leslie Kish International Fellows Fund, contact Patrick Shields, Development Director, Institute for Social Research, University of Michigan, 426 Thompson Street, Ann Arbor, Michigan 48109, USA, or send an email to peshield@umich.edu.)



"Keep young by being curious, and have young friends"



SOFTWARE REVIEW



g-DESIGN AND g-CALIB-S: SPSS® MODULES FOR GENERALISED CALIBRATION

Camille Vanderhoeft,
Jean-Marc Museux, Etienne Waeytens
Statistics Belgium

1. Introduction

Generalised calibration (Deville et al., 1993) is a popular and powerful framework in survey estimation. It allows using different sources with auxiliary information to improve estimates from sample surveys: sampling error can be reduced, and more precise estimators are obtained. It is also believed that generalised calibration can correct appropriately for nonresponse.

Statistics Belgium traditionally applied poststratification techniques, with all its limitations, for grossing up survey data. There is now a strong tendency at Statistics Belgium to move to more sophisticated calibration techniques: we are looking for more appropriate auxiliary information to be used for estimation from the Labour Force Survey, Household Budget Survey, Time Use Survey, and Tourism Survey, as well as from several business surveys. As more external data sources become available and are being exploited more efficiently, there is an increasing need in appropriate estimation tools at Statistics Belgium. Since about four years now, SPSS is the main tool for our statisticians to calculate results, to produce tables, and to carry out statistical analyses after basic databases are prepared by the information department. Up to now, no SPSS module existed for generalised calibration. Therefore, we decided to develop *g-DESIGN* and *g-CALIB-S* under SPSS 9.0, which make extensive use of SPSS's syntax language, matrix language, and macro facilities.

Our modules are comparable to Calmar, developed as a SAS module at INSEE, France (Sautory, 1993). With *g-CALIB-S* virtually any calibration model can be applied. More specifically, we are not restricted to calibration on margins (or totals of quantitative variables within categories of qualitative variables). The price for this generality, on the other hand, is that preparation of the input files for *g-CALIB-S* can be very complicated and is, therefore, not easy to automate. However, *g-DESIGN* contains several macros that are supporting to a large extent the construction of these files. We demonstrate this later in Section 3.1. Our current experience shows that those macros are indeed the core of the procedure to transform basic survey, as well as population data, into the required input files for the calibration module.

This article is primarily intended to introduce the statistical world to *g-DESIGN* and *g-CALIB-S* and to show how an efficient tool for calibration has been created under SPSS. The text is not giving the mathematics of the underlying properties of the general method used for solving the calibration problem. A more lengthy discussion is given in Vanderhoeft et al., (2000).

2. The Calibration Problem

Consider a sample *s* of size *n* from a population *U* of size *N*. Let the *n* sample elements be selected according to a sampling design with positive inclusion probabilities $\pi_1,\ldots,\pi_N$ for all elements in *U*. Suppose that measurements on *m auxiliary* or *calibration variables* are available for all sample elements; let x_{kj} be the

value of the *j*-th calibration variable for the *k*-th sample element. The $n \times m$ matrix $\mathbf{X} = \left(x_{kj}\right)$ is called the (*calibration*) *design matrix*. It is assumed that qualitative variables are already transformed into sets of indicator variables, etc. Finally, the population or *calibration total* for each of the calibration variables must be available. The generalised calibration problem consists of adjusting the *initial* or *sampling weights* $d_k = 1/\pi_k$, resulting in *adjusted* or *calibrated weights* $w_k = g_k d_k$, where the g_k are the adjustment factors or *g-weights*. The calibrated estimator of the total of *y,* an arbitrary variable of interest, is a Horvitz-Thompson-like estimator using the calibrated weights instead of the sampling weights: $\hat{t}_y^{CAL} = \sum_{k=1}^{n} w_k y_k$.

The generalised calibration problem can be formulated as a non-linear optimisation problem as follows.

(C1) Minimise the *distance* $\sum_{k=1}^{n} d_k G\left(g_k\right)$ as a function of g_k ,

(C2) subject to *m calibration constraints* $\hat{t}_j^{CAL} \equiv \sum_{k=1}^{n} d_k g_k x_{kj} = t_j \ (j = 1, \ldots, m)$,

(C3) and, occasionally, subject to *boundary constraints* $L \le g_k \le U \ (k = 1, \ldots, n)$, with $L < 1$ and $U > 1$.

The so-called *distance function G* is measuring the difference between the g-weights g_k and 1. It can be chosen conveniently (Deville et al., 1993), considering practical properties of the resulting g-weights. Often a quadratic function is used (i.e., $G(x) = \frac{1}{2}(x-1)^2$); the calibration method is then said to be *linear*. Estimators based on this method are called GREG estimators. Other choices of *G* may force the calibrated weights being positive. Under the linear method, the additional constraints (C3) can be used to restrict the g-weights. The so-called *logit* method implicitly works with such boundary constraints.

The calibration constraints can be written in matrix form as $\mathbf{X}^T \mathbf{w} = \mathbf{t}$, where $\mathbf{t} = \left(t_1, \ldots, t_m\right)^T$ is the *m*-vector of population or *calibration totals* and $\mathbf{w} = \left(w_1 = g_1 d_1, \ldots, w_n = g_n d_n\right)^T$ is the *n*-vector of calibrated weights. Finally, we shall say that the calibration design matrix is a formal representation of a *calibration model*.

The solution of this non-linear optimisation problem can be obtained by an iterative procedure (Deville et al., 1993; Vanderhoeft et al., 2000). Each iteration requires the inversion of the $m \times m$ matrix of weighted sums of squares and cross products of the observed calibration variables (i.e., the weighted SSP-matrix). In each iteration, the weights in those sums are updated estimates of the calibrated weights. The use of the ordinary inverse of the weighted SSP-matrix is only possible if the design matrix $\mathbf{X}$ has a full rank *m*. So, in constructing this matrix, some care should be taken in order to avoid any linear dependency between the observed calibration variables. This can be very difficult to tackle when the number of calibration variables is large due to the complexity of the calibration model.

Our implementation in SPSS uses, instead of the ordinary inverse matrix, a *generalised* or *g-inverse*. A g-inverse of a non full rank matrix always exists and consequently the iterative procedure will not be stopped due to a non full rank design matrix. The use of g-inverses requires some mathematical justification, since the g-inverse of a given matrix need not be unique. This mathematical justification can be found in Vanderhoeft et al., (2000). They show that the properties of the iterative procedure (convergence rate, the solution, …) are invariant for the choice of the g-inverse.

In our module *g-CALIB-S*, we use the Moore-Penrose inverse, which is available in the matrix language of SPSS 9.0. Most packages that include matrix manipulation allow to calculate g-inverses. SAS/IML is another example.

From a practical point of view it is interesting to be able to work with a design matrix with linear dependencies between the columns. It simplifies either the preparation of these matrices (if the software does not construct it automatically) or the implementation of its construction from originally observed variables. Of course, the construction of a design matrix is not unique for a given calibration model. *Equivalent* design matrices (i.e., matrices that generate the same linear vector space or have the same projection matrices), however, will lead to the same solution and do not affect the iterative procedure.

3. SPSS Implementation

3.1 The Module *g-DESIGN* and Input Files for *g-CALIB-S*

The module *g-CALIB-S* needs two input files, called `Cal_SURVEY.sav` and `Cal_TOTALS.sav` hereafter. We now discuss the contents and structure of these files and how they can be constructed from original (raw) data files using macros in *g-DESIGN*. `Cal_SURVEY.sav` should finally contain the transformed survey data (i.e., the calibration design matrix **X**), together with the initial weights (often the sampling weights). `Cal_TOTALS.sav` should finally contain the corresponding calibration totals (i.e., the vector **t**.)

`Cal_SURVEY.sav`: The survey data input file

We illustrate the transformation of survey data into one of the input files for *g-CALIB-S* by means of a small hypothetical example. Suppose that the original SPSS survey data file contains data on 8 cases, in 5 variables or columns, as follows:

```
   ID WEIGHT STRATUM     A      B      Z
1.00  10.00    1.00  1.00   1.00  15.00
2.00  10.00    1.00  1.00   1.00  17.00
3.00  12.00    1.00  1.00   2.00  12.00
4.00  10.00    1.00  1.00   2.00  12.80
5.00   5.00    1.00  2.00   1.00  10.50
6.00   5.00    1.00  2.00   1.00  12.00
7.00   5.00    1.00  2.00   2.00  13.40
8.00   5.00    1.00  2.00   2.00  12.00
```

The column headings are the variable names, stored in the data dictionary of the survey data file. We assume that A and B are qualitative variables, while Z is a quantitative variable. STRATUM is qualitative too, with, in general, *L* categories. It is important for this variable that all categories 1, 2, …, *L* are represented in the survey data file. In our example, $L = 1$. STRATUM essentially is a calibration variable too, but its purpose is different from the other qualitative calibration variables A and B. This will be clarified later. In our applications at Statistics Belgium, STRATUM often determines geographical areas, such as regions, provinces, etc.

Standard SPSS matrix commands are used to store the case identifiers, the case weights, STRATUM, and the survey variables A, B and Z into vectors `ID`, `WE`, `STRATUM`, `A`, `B` and `Z`, say. (Names are not really important, so far.) Now, suppose that the maximal calibration model contains the terms A, B, Z, A.B, A.Z, B.Z, and a constant term 1. Then *g-DESIGN* provides the required macros `DesC1`, `DesC2`, and `DesC1Z` to transform the vectors `A`, `B`, and `Z` into the corresponding columns of the (maximal) calibration design matrix **X**. The constant term is easy to construct, using matrix functions `nrows` and `make`. The matrix commands are as follows:

```
compute X0 = make(nrows(ID),1,1)            /*  Term: 1     */.
DesC1 var=A des=XA lab=LabA                 /*  Term: A     */.
DesC1 var=B des=XB lab=LabB                 /*  Term: B     */.
DesC2 var1=A var2=B des=XAB lab=LabAB p=1   /*  Term: A.B   */.
DesC1Z var=A zet=Z des=XAZ lab=LabAZ        /*  Term: A.Z   */.
DesC1Z var=B zet=Z des=XBZ lab=LabBZ        /*  Term: B.Z   */.
```

`X0`, `XA`, `XB`, `XAB`, `XAZ`, `XBZ`, and also `Z` are matrices, with, in our example, numbers of columns equal to 1, 2, 2, 4, 2, 2, and 1, respectively. Those matrices are the corresponding parts of the design matrix **X**. *g-DESIGN*

contains three other macros, DesC3, DesC2Z, and DesC3Z, which can be used to construct parts of the design matrix, corresponding to higher order interaction terms in the calibration model (if more qualitative variables are available). LabA, LabB, … are row vectors containing appropriate labels for the columns of XA, XB, ….

The survey data input file Cal_SURVEY.sav is then easily created by the following SPSS matrix command, which saves the original data and the (maximal) calibration design matrix into that file:

```
save {ID, WE, STRATUM, A, B, Z, X0, XA, XB, XAB, XAZ, XBZ}
     /outfile = 'C:\my documents\calibration\Cal_SURVEY.sav'
     /variables = CASE,WEIGHT,STRATUM,A,B,Z,X0,A1,A2,B1,B2,
                  AB11,AB12,AB21,AB22,A1Z,A2Z,B1Z,B2Z.
```

The file Cal_SURVEY.sav now contains the following data:

CASE	WEIGHT	STRATUM	A	B	Z	X0	A1	A2	B1	B2	AB11	AB12	AB21	AB22	A1Z	A2Z	B1Z	B2Z
1	10	1	1	1	15	1	1	0	1	0	1	0	0	0	15	0	15	0
2	10	1	1	1	17	1	1	0	1	0	1	0	0	0	17	0	17	0
3	12	1	1	2	12	1	1	0	0	1	0	1	0	0	12	0	0	12
4	10	1	1	2	13	1	1	0	0	1	0	1	0	0	13	0	0	13
5	5	1	2	1	11	1	0	1	1	0	0	0	1	0	0	11	11	0
6	5	1	2	1	12	1	0	1	1	0	0	0	1	0	0	12	12	0
7	5	1	2	2	13	1	0	1	0	1	0	0	0	1	0	13	0	13
8	5	1	2	2	12	1	0	1	0	1	0	0	0	1	0	12	0	12

The top row is just the list of variable names – as defined by the user – stored in the data dictionary of the file. The last 14 columns, from Z to B2Z, constitute the (maximal) design matrix **X**. Notice that *z*-values are displayed as rounded to integer values; the exact values are, however, stored in the file. It is clear that there are several linear dependencies between the columns: the matrix **X** has 14 columns, but its rank is only 7.

As it was said before, the variable STRATUM has a very special purpose and must be included in Cal_SURVEY.sav. In our hypothetical example, its value is 1 for each case. This means that only one stratum is considered. When several strata exist, an independent calibration is done in each stratum according to the model defined. An important consequence is that the data can be loaded and used to calibrate, stratum by stratum, which avoids exceeding memory limits.

Cal_TOTALS.sav: The totals input file

In our hypothetical example, the following variables in Cal_SURVEY.sav must also be present in Cal_TOTALS.sav: STRATUM, Z, X0, A1, A2, B1, B2, AB11, AB12, AB21, AB22, A1Z, A2Z, B1Z, B2Z. Again in our example, since STRATUM assumes only the value 1, the file Cal_TOTALS.sav has only 1 record, containing the appropriate calibration totals. In general, there is one record for each calibration stratum, holding the subpopulation totals for all calibration variables in each stratum. In our hypothetical example, we assume that, from an external source, or from another survey, the following (estimated) totals could be obtained:

STRATUM	Z	X0	A1	A2	B1	B2	AB11	AB12	AB21	AB22	A1Z	A2Z	B1Z	B2Z
1	1190	90	60	30	44	46	29	31	15	15	850	340	620	570

These totals are fully consistent (e.g., 1190 = 850 + 340 = 620 + 570). The total 90, corresponding to the constant variable X0, can be interpreted as the size of the stratum 1 in the population.

The two input files for our hypothetical example are now ready. The user should always remind that the name of the calibration stratum variable, in both files, must always be STRATUM, and that the case identifier and weight variable in Cal_SURVEY.sav must be CASE and WEIGHT, respectively. Names of other variables are arbitrary, as far as general SPSS rules are satisfied. The order of appearance of the 14 potential calibration variables Z, X0, … B2Z is also arbitrary, but it must be the same in the two files.

3.2 The Module *g-CALIB-S*

We now discuss some features of the module *g-CALIB-S*. A final calibration model will be specified for our hypothetical data, and part of the output is shown.

First of all, an auxiliary module *g-PREPARE* has to be run before *g-CALIB-S* is executed. This auxiliary module sorts the survey data by STRATUM. It also counts the number of cases in the survey data file for each calibration stratum and computes the first and last record number for each stratum in the sorted survey data file. This summary information is added to the data in the totals file, which is also sorted by STRATUM. Both sorted files are stored again, using the same names and are the final input files for *g-CALIB-S*.

Both *g-PREPARE* and *g_CALIB-S* are activated through a SPSS Production Facility job. When the job is started, a dialogue box with a list of user prompts appears. The user can change the default answers for these prompts. SPSS transforms the user's answers into macros, with internal names `@WORKDIR`, `@XDATA`, etc. They are used as parameters in our programs. For our illustrative example, we enter the following:

(i) The working directory and the input files:
- @WORKDIR = C:\My Documents\Calibration\
- @XDATA = Cal_SURVEY.sav
- @CALTOT = Cal_TOTALS.sav

(ii) A list with the names of the calibration variables finally being used:
- @XVARS = Z,X0,A1 to B2 (*the order in which the selected calibration variables are listed may differ from the order in the input files*)

(iii) The strata for which calibration has to be performed:
- @STR_1 = 1
- @STR_N = 1

(iv) The calibration method (1 = linear method, 2 = raking ratio method, 3 = truncated linear method, 4 = logit method):
- @TYPE = 1

(v) The positive value of an overall scaling factor, or 0 or any negative value if the program has to calculate it before calibration actually starts:
- @SCALE = 0

(vi) Lower and upper bounds on the g-weights for methods 3 and 4:
- @L = 0.7 (*not used here, since the linear method is chosen; default value*)
- @U = 1.5 (*not used here, since the linear method is chosen; default value*)

(vii) Iteration parameters (tolerance and maximum number of iterations):
- @TOL = 0.001 (*default value*)
- @ITERMAX = 20 (*default value*)

(viii) A logical parameter managing output (Y if, apart from the basic calibration results, intermediate results are required; N otherwise):
- @INFO = N

Most of these are easy to understand. The scaling factor in (v), however, deserves some special attention.

Let ϕ denote the scaling factor. The initial weights, which are usually the sampling weights $d_k = 1/\pi_k$, are multiplied by ϕ. If ϕ is specified by the user, the same fixed value is applied for all calibration strata. If ϕ is to be calculated by the program, the value 0 or any negative value must be entered at the user prompt (as we did in the example); the value for ϕ will then be calculated for each stratum separately. This parameter can act as a preliminary and overall up-weighting factor to produce a first correction for nonresponse. The inclusion of the scaling factor ϕ is necessary to improve the numerical performance and reliability of our calibration program, especially for the methods with truncation of the g-weights.

For each stratum, the output file of *g-CALIB-S* shows general information on the data, some details about the method and the iteration, and the computed or fixed value of the scale parameter. Next, the following statistical results are shown: the initially fixed calibration totals; the initially calculated estimates (usually the

Horvitz-Thompson estimates), based on the calibration variables and the initial weights in the survey data file; scaled estimates, obtained by multiplication of these initial estimates with the scale parameter ϕ; the final calibration estimates (if the procedure converges); and the relative difference between the latter and the fixed totals. For our example, this part of the output looks as follows:

```
FIXED, INITIAL, SCALED and CALIBRATED TOTALS per aux. variable :
           Fixed       INI_est       SCA_est       CAL_est        % DIFF
Z    1190.000000    831.500000   1190.000000   1190.000000       .000000
X0     90.000000     62.000000     88.731209     90.000000       .000000
A1     60.000000     42.000000     60.108238     60.000000       .000000
A2     30.000000     20.000000     28.622971     30.000000       .000000
B1     44.000000     30.000000     42.934456     44.000000       .000000
B2     46.000000     32.000000     45.796753     46.000000       .000000
```

The relative difference may help identifying problems with the data, such as an empty category of a variable with corresponding non-zero fixed total, inconsistencies with respect to the fixed totals, etc. Notice that the scale parameter ϕ is computed from two totals corresponding to the first calibration variable (i.e., Z, in the list: 1190/831.5 = 1.4331148527). If X0 would be the first variable in the list, the value of ϕ would be 90/62 = 1.4516. The reciprocal of this value, 0.6889, can be interpreted as an estimated global nonresponse rate.

Finally, the output contains summary measures describing the distribution within the strata for the three weight variables and corresponding boxplots. We show the parameters for the distribution of the g-weights in the above example:

Distribution of G-WEIGHTS within strata

			Minimum	Median	Mean	Maximum	Std Deviation
STRATUM	1.00	G_WEIG	.82	1.02	1.02	1.28	.16

			Sum	Percentile 05	Percentile 25	Percentile 75	Percentile 95
STRATUM	1.00	G_WEIG	8.17	.82	.87	1.13	1.28

A SPSS data file is produced too, holding for each case *k* the scale parameter ϕ (variable SCALE), the scaled weight $\phi \cdot d_k$ (variable SCAWEI), the calibrated weight w_k (variable CALWEI) and the g-weight $g_k = w_k / (\phi \cdot d_k)$ (variable G_WEIG), together with the identification of the case (variable CASE) and the stratum to which the case belongs (variable STRATUM). This file can easily be merged with the original survey data file, allowing routine calculation of estimated totals for survey variables.

We have applied a calibration model that can be summarised as 1 + Z + A + B, meaning that we have calibrated on the total size of the population (in stratum 1), the marginal population size in categories of A and categories of B, and the total of the quantitative variable Z in the population. The resulting g-weights and calibrated weights are all positive. If we would have used all calibration variables that were prepared in the input files, then we would have obtained some negative weights. The user can try this himself.

4. Final Remarks and Conclusion

The reader may by now have understood that "*g-*" in *g-CALIB-S* stands for "*generalised*" in "generalised calibration", but at the same time also refers to the use of *g-inverses* of matrices. "*–S*" refers to the inclusion of the calibration *stratum* variable. The name *g-DESIGN* has a similar explanation.

We have developed a flexible tool for generalised calibration estimation of totals from sample surveys. The strength of our program is a direct consequence of 1° using g-inverse matrices and 2° the completely general framework, which allows to apply any possible calibration model. We are aware of the fact that our program

still has some minor weaknesses, but most of these have to do with failure of the program – currently without a clear error message – if the data files have not correctly been constructed. So the user can easily avoid problems of that nature by setting up the data carefully.

More important, Statistics Belgium now has a tool that can be run under the statistical system that has been opted for a few years ago: SPSS for PCs. We already have the feeling that our statisticians are encouraged to look for and use more auxiliary data in external sources. A tool such as *g-CALIB-S* can only positively contribute to this tendency.

We have developed *g-DESIGN* and *g-CALIB-S* under SPSS 9.0 for Windows Base; no special modules are required in that version. We have tested, with success, the new modules also under SPSS 8.0, extended with the Advanced Statistics module that provides the matrix language. Finally, we regret to have to warn the user for running the modules under SPSS 10.0: some matrix commands were causing serious trouble. We did not try to solve the problems by rewriting parts of the modules, since we feel that our programming logic should be correct, and that the problems are caused by bugs in the basic software. So the user is recommended simply to run the calibration modules only in SPSS 9.0 or 8.0. Older versions have not been tested.

5. Delivering the Software

The modules can be obtained from the authors at the following address:

Statistics Belgium
Methodology and Co-ordination
Rue de Louvain, 44
B-1000 Brussels
Belgium
Tel: +32 2 548.66.44, Fax: +32 2 548.63.76
E-mail: camille.vanderhoeft@statbel.mineco.fgov.be

The software is delivered with some example files. This text, as well as a text containing the mathematical justification of our approach to solve the calibration problem (Vanderhoeft et al., 2000), can be sent too (in Word format). All files can be sent as attachments with an email message. The same address can also be used to ask for support.

References

Deville, J.-C., Särndal, C.-E. and Sautory, O. (1993). Generalized Raking Procedures in Survey Sampling, *JASA*, Vol. 88, No. 423, pp.1013-1020.

Sautory, O. (1993). La macro SAS CALMAR: Redressement d'un Echantillon par Calage sur Marges, *Document de travail de la Direction des Statistiques Démographiques et Sociales*, no. F9310.

SPSS Inc. (1999). *SPSS® Base 9.0 Syntax Reference Guide*, SPSS Inc.

Vanderhoeft, C., Waeytens, E. and Museux J.-M. (2000). Generalised Calibration with SPSS® 9.0 for Windows Base. Paper presented at the *2ème Colloque Francophone sur les Sondages*, Brussels. To appear in the proceedings.

BELGIUM

by Etienne Waeytens, Camille Vanderhoeft, and Jean-Marc Museux

Calibration at Statistics Belgium: *Statistics Belgium* has put a lot of energy into the development of SPSS modules for generalised calibration. Two modules, g-DESIGN for preparation of the data, and g-CALIB-S for calculation of calibrated weights, are now ready. The modules are freely distributed in the statistical world. At *Statistics Belgium*, they are being integrated in systems for various surveys. The modules are fully general, and do not put any restriction on the generalised calibration methods and models as developed by J.C. Deville, C.E. Särndal, and others.

The availability of powerful tools for calibration encourages our statisticians to look for more and better external data, in order to improve the quality of the estimates. Classical post-stratification based weighting will be replaced step by step by other calibration techniques. It is also a step forward towards systematic calculation of variances of estimators, and consequently a more efficient evaluation and management of the quality of our surveys.

Contact person:
camille.vanderhoeft@statbel.mineco.fgov.be.

Use of administrative information – reduction in burden on respondents: The INS has embarked upon an extensive program to simplify administration and reduce the burden on respondents. The systematic use of administrative sources is encouraged to replace or reduce direct collection of data. Several ways of using administrative data are contemplated:

1. Use of administrative information as a substitute for surveys. Thus, the annual Salary Structure and Distribution Survey uses the information available as much as possible. A simplified questionnaire is sent to companies to obtain information that is not available in the records. The next Labour Cost Survey 2000 will be based essentially on administrative records. Information that is not available there will be obtained from a reduced sample of companies. Administrative information collected earlier will be used to calibrate the estimators.
2. Administrative information is used as supplementary information in calibrating estimates. The resulting increase in accuracy would make it possible to consider a reduced sample size. This method will be used in the Labour Force Survey.

Contact person:
annie.versonnen@statbel.mineco.fgov.be.

Labour Force Survey will now be conducted on an on-going basis: Since January 1999, the labour force survey (Enquête sur les Forces de Travail) has been conducted on an on-going basis. Every week approximately 920 households are surveyed, initially in person. A second survey is conducted by telephone and/or by mail 3 months later. At the present time, quarterly estimates are provided for the main indicators within 12 weeks. In the case of the detailed indicators, only annual averages are produced by aggregating the data over one year. The redesign of the survey coincides with the launch of an extensive project designed to improve the quality of the results of the survey. Generalized calibration techniques are used to deal with the non-response problem and to improve estimates. In addition, household and individual statistics, among others, will become more consistent. The use of administrative information concerning the social security of individuals will also make it possible to ensure greater consistency between the results of the survey and administrative statistics. Calibration techniques will also make it possible to take into account the temporal aspects of the survey and to combine the results of the first and second surveys. These new techniques should be applied to the data from the 2000 survey.

Contact person:
denis.luminet@statbel.mineco.fgov.be,
jeanmarc.museux@statbel.mineco.fgov.be.

Changes in producer prices - new telephone survey: To meet specific needs such as changes in input prices in the construction sector, a new survey of the price of goods and services has been instituted. To reduce the burden on respondents and to minimize production times, the survey is conducted by telephone. A special unit (the Call Centre) has been created. Interviews (lasting approximately one minute) are computer-assisted (CATI). The software used is BLAISE. The survey field will expand gradually and by early 2001 will cover all manufacturing products. In this segment, products and producers are sampled by means of a random sampling algorithm proportionate to the value of the transactions concerned. Then, an extension is included to cover services and the economy as a whole. The results obtained can be used to create a high-quality monthly producer price index.

For further information, contact:
marc.debusschere@statbel.mineco.fgov.be
or: http://statbel.fgov.be/indicators.

CANADA

by Gordon Brackstone

Statistics Canada has published the results of its first economy-wide survey on electronic commerce and business use of the Internet. The Information and Communications Technologies and Electronic Commerce (ICTEC) Survey collected information by mail from a sample of some 23,000 businesses between October 1999 and March 2000, for reference year 1999. The sample design, which used the sample from an existing survey of businesses, covered all industries except construction, local government, production of crops and animals, fishing, hunting and trapping. Items such as the use of computers, e-mail and the World Wide Web were covered, as well as actual dollar values for sales over the Internet. Data were published by industry sector, as well as by public or private sector. The results are available in the August 10, 2000 edition of *The Daily* at the Statistics Canada web site (www.statcan.ca). Plans are now underway to repeat the survey in a similar format for reference year 2000.

For further information, contact Patricia Whitridge, Statistics Canada, tel.: (613) 951-1475, or by e-mail at whitpat@statcan.ca.

The National Graduates Survey (NGS) and follow-up report are both parts of a longitudinal survey of individuals who have completed a post-secondary program during a given year. The main purpose of the survey is to examine the labour market experiences of recent graduates and how they financed their studies.

The reference years used for the past four editions of the survey are 1982, 1986, 1990 and 1995; the next edition should cover reference year 2000. Students who successfully fulfilled the requirements for a college or university degree, diploma or trade/vocational certificate during the reference year constitute the target population. Based on administrative data from provincial and territorial governments, approximately 20% of this population (about 60,000 out of 300,000) is sampled by means of a design stratified by province, level of education and field of study. Given the relatively small population size and usually low response rate of doctoral-degree holders, the latter are enumerated by census. The NGS is conducted two years after diplomas are received; the follow-up report on NGS respondents is made three years later, or five years after diplomas are received. Thus, 1995 graduates were surveyed in 1997 (NGS) and 2000 (follow-up report). With the help of computers, data are collected by telephone at Statistics Canada Regional Offices.

There are two main methodological challenges inherent in the survey. First, the creation of a survey frame is a long-range undertaking that requires considerable human resources; in the future, a standardized system will be implemented that should make this step easier and less burdensome. Second, as the target population is relatively young and mobile (for the 1997 NGS, the median age was 26), tracing the individuals sampled represents a key factor in the survey's success, both at the NGS and follow-up levels.

For further information, contact Jacques Taillon (Centre for Education Statistics, Statistics Canada (613) 951-3332, jacques.taillon@statcan.ca) or Pierre David (Social Survey Methods Division, Statistics Canada, (613) 951-4345, pierre.david@statcan.ca).

CÔTE D'IVOIRE

by Benjamin Zanou

The National Graduate School of Applied Economics and Statistics (*École Nationale Supérieure de Statistique et d'Économie Appliquée – ENSEA*) prepared and conducted two surveys during the second and third semesters of 2000:

- a Multi-Indicator Cluster Survey - MICS;
- a Survey on the Use of Medical Services in Côte d'Ivoire.

The purpose of the MICS is to collect data and use it to depict the status of children and women in Côte d'Ivoire. It follows the involvement of member states at the World Summit for Children held in New York in 1990. The MICS 2000 followed the first MICS which was conducted in 1996 to measure the extent to which Côte d'Ivoire achieved its mid-decade objectives.

The MICS is a nation-wide sample survey covering 292 clusters (census enumeration areas), each comprising 1,000 individuals. Following the inventory of cluster households, 25 households were drawn at random to participate in the survey.

The collection material includes three questionnaires (household, women and children under 5). The field activities were carried out by 36 officers and 14 supervisors over the course of two months.

Data capture took place subsequent to the field activities and the data files are currently being cleaned. The results will be used to measure the achievement of goals. UNICEF financed the study on behalf of the Government of Côte d'Ivoire.

The purpose of the Survey on the Use of Medical Services in Côte d'Ivoire is to identify factors that limit the use of medical services. Its random sample contained 90 clusters (30 urban and 60 rural), selected from 81 localities according to their size. Clusters are spread out in 7 out of the 10 administrative regions of the former zoning. From each first-level cluster, three sections are drawn by judgement sampling. These sections were used to select the households to be interviewed, i.e., 11 households from rural clusters and 22 households from urban clusters. A total of 1,291 households were involved in the survey.

The collection material includes a multi-component questionnaire and a fact sheet. It took about 40 days for 25 survey officers to carry out the collection. Now that field activities and data capture are done, data analysis is being carried out.

For more information, please contact Benjamin Zanou (bzanou@ensea.ed.ci).

ESTONIA

by Imbi Traat

Since January 2000 the Statistical Office of Estonia has run its **Household Budget Survey** (HBS) according to a redesigned methodology. The redesign works, financially supported by the World Bank, lasted nearly two years. Four working groups handled different aspects of the survey. As a result many changes were introduced into the previous survey, launched in 1995. The author of the present report lead the working group responsible for the sampling and estimation methods. There were two major shortcomings in these methods – the incorrect inclusion probabilities due to household replacements and biases in the estimates due to nonresponse.

From the previous survey it was known that about 14 % of the households do not live at the addresses sampled from the Population Register. Due to lack of time the household actually living at the address was included in the sample (replacement household) and the selected one from the Register was not traced further. The inclusion probability was formed on the basis of the replacement household size. The true inclusion probability, in fact, is determined by the number of traces leading to that address, i.e., by the number of people having this address in the Register. Careful study of the Population Register showed that it is possible to find inclusion probabilities of the addresses on the basis of the Register. In the new survey a stratified systematic sample is selected from the Population Register, so that data on each sampled record includes in addition to the name and address identifications also the frequency of the corresponding address in the Register. If the frequency is less than or equal to seven the household is included according to the address-rule (no matter which household lives at that address), and the frequency is used for calculation of its inclusion probability. If the frequency is greater than seven the household is included according to

the person-rule, meaning that the selected person has to be found and his/her household taken. The inclusion probability in this case will be formed on the basis of the household size given by the household. The analysis showed that the addresses having above frequency seven in the Register were usually incomplete (e.g., consisting only of a village name in rural areas), but good enough to apply person-rule. The analysis and the pilot survey confirmed that about 86% of the sampled households can be located by the address-rule and 14% by the person-rule.

In the estimation stage regional poststratification was changed to improve estimates in the presence of varying response rates and income-expenditure levels over the regions. Calibration on the population sex-age distribution was added. Major estimates are published with standard errors. The package SUDAAN is used for calculations. The redesign works are documented in Traat, I. (1999) Redesign of the Estonian Household Budget Survey. *Final Report of the Sampling Group.* Tallinn: The Statistical Office of Estonia. A paper for the journal *Statistics in Transition* is under preparation.

The Statistical Office of Estonia performed its **Population Census** in March 31 – April 9, 2000. The Census was carried through in the traditional way, by interviewers visiting each living place and filling in Personal and Housing Questionnaires. As technological innovation the completed questionnaires were scanned into the computer and processed by the software for optical detection.

The preliminary data of the Census 2000 have been published in the monthly bulletin *Estonian Statistics, 2000, No. 6*. Accordingly, the population *de facto* was 1,370,500 (all persons in Estonia at the Census moment March 31, except the diplomatic staff and their household members). The usually resident population of Estonia was 1,376,700 (persons with permanent place of residence in Estonia including those temporarily (for up to one year) in foreign countries). Both numbers have greatly (about 200,000) decreased compared to the 1989 Census. The decrease is mostly caused by people in the service of the army of the former Soviet Union and their family members who have left Estonia, but also by the negative natural increase. The undercoverage of the Census is preliminarily estimated to be 2%. The final Census results will be completed by stages in 2001-2002.

The **Workshop on Survey Sampling Theory and Methodology** for Baltic and Nordic countries on June 18-22, 2000 was already the fourth in a series. This time it took place in the small town Pärnu of Estonia. It was organized under the auspices of the Institute of Mathematics and Informatics (Lithuania), the University of Latvia, the University of Tartu (Estonia), Umeå University (Sweden), Vilnius University (Lithuania), the Central Statistical Bureau of Latvia, Statistics Lithuania and the Statistical Office of Estonia. The series of workshops has been initiated by Professor Gunnar Kulldorff (Sweden) who is also a member of the Organizing Committee. The majority of the financial support came from the Swedish Institute.

The workshop continued to realize the same aims as its predecessors – to educate the young sampling specialists, to enhance their theoretical and practical skills, to strengthen contacts and cooperation among the Baltic and Nordic sampling people. Participants were teachers and research students of the universities and young statisticians of the statistical agencies of these countries (22 from Estonia, Latvia and Lithuania, and 8 from Finland and Sweden). The main lecturer was Professor Lennart Bondesson (Sweden). The first topic of his four-hour lectures was area sampling, a methodology for sampling of populations spread over a geographical region with its main application in forest inventory. The second topic was more theoretical introducing a new sampling method based on renewal theory, being a generalization of systematic sampling by allowing a random step length. There were eight lectures by other scholars and nineteen contributions of the young participants. Contributions were coupled with discussions which was a fruitful tutorial component of the workshop.

FINLAND

by Kari Djerf

Statistics Finland's new Main Lines of Research and Development in 2000-2003 was published in August. As a continuation of a similar program for the years 1997-1999, this edition covers the focal points of research activities in social sciences, economics and statistical science orientation. One of the key issues is fostering scientific networking

with both local and international universities, statistical agencies and research institutes. For more information, please contact Mr. Timo Byckling at Statistics Finland (e-mail timo.byckling@stat.fi).

The Health 2000 Survey, one of the largest surveys ever in Finland, was started in mid-September. It is a joint effort of various partners, the most important of which are the National Public Health Institute (coordinator), Central Pension Security Institute, Finnish Institute of Occupational Health, National Research and Development Centre for Welfare and Health, Social Insurance Institution and Statistics Finland. Additionally, numerous medical researchers from various universities have been actively planning the survey contents of their special fields.

The survey covers practically all health aspects. Data collection begins with a home-interview after which all respondents over 30 years are called for a one-day health examination which takes place in the designated health care centres or university hospitals. Direct data collection will last until March 2001 after which various health-related administrative information is merged to the basic data during the next years, provided that the respondent gives his or her written permission.

The sampling design in this survey is a two-stage stratified PPS design where the PSUs are health centre regions which may consist of one or more municipalities. The PSUs were stratified according to the geographic boundaries of five university hospitals and by the size. First, the 15 largest cities were selected with probability 1, and after that the population of the remaining health care centre regions was used as the measure of size in selecting the other PSUs. We have altogether 80 PSUs and the total sample size is 10,000. That sample is further divided into two different study contents according to age: 8,000 individuals aged 30 years or more to participate both in the interview and the clinical examination, and 2,000 individuals between 18 and 29 years to participate in the interview part only.

The survey is a continuation of the Mini-Finland survey conducted in the late 1970s. We expect this data will be used as a basic reference for the next 20 years in the public health studies. For more information, please contact Mr. Kari Djerf at Statistics Finland (kari.djerf@stat.fi).

FRANCE

by Benoît Riandey

Assessment of the 1999 French Census: Population registers were part of the French Revolution's legacy to Western Europe, but ironically they have never worked in France, and the French authorities responsible for the protection of privacy would not allow their establishment today.

On the other hand, the people of France are very receptive to the French census, because it produces information that is vital to management and finances at the local level. The rate of coverage is high. The control survey for the 1990 census assessed omissions at 1.8%, and double counting at 0.7%. In 1999, the latter seems to be even lower, and omissions may be slightly higher, which is generating some doubts about net migration figures.

As in the past, the survey was carried out by an army of census-takers (there were 115,000 in 1999) recruited by the *communes,* or municipalities, and was organized on a deposit-withdrawal basis under the technical supervision of the statistical institute. The questionnaires issued to the census-takers – one double-sided sheet of A4 paper per person – are stored and monitored in the local mayor's office, but municipal officials have no access to them. This circumstance prompts some reluctance among the respondents, although there are no questions on such sensitive issues as income. The number of rooms in a dwelling is felt to be the most sensitive question asked, because of an unwarranted fear that there may be tax repercussions.

Two recent changes in the census are of interest. The names and addresses of students were entered provisionally and sorted both to avoid double counting and to reduce the risk of omitting a household. Supervisors compare the number of households counted at any one address with the numbers provided by the income tax service. Optical reading of the questionnaires was the major technical innovation of this census, and used the experience gained through the Israeli census. The difficulty lies not so much in the optical reading and character recognition, as in the need to process a fairly large volume of questionnaires (that is, "images") in a relatively short time. The second technical innovation in 1999 was the use of

automatic coding for various identifiers (*commune*, nationality, social category, economic sector, and so on) based on inputted data, without manual processing.

Since 1954, INSEE (France's national economic research and statistics institute) has combined with the census a survey of families using a self-administered questionnaire. In 1999, it covered some 400,000 men or women aged 18 or older, or one district in 50. The questionnaire – on a double-sided sheet of A3 paper – covered previous unions, whether by marriage or otherwise, all births and adoptions, and the passing on of regional languages. The sensitivity of such questions mandated an optional survey, to be returned directly to INSEE (francois.clanche@insee.fr). The Survey Statistician, in issue No. 39, described a Disability Incapacity Dependency Survey using an 18-question filtered questionnaire combined with the census in one district in 150. The results were fully satisfactory (pierre.mormiche@insee.fr).

The conditions for distribution of local statistics (those produced within each *commune*) generated considerable debate: geographers wanted access to high-resolution microdata and tables on smaller districts than the standard districts of 2,000 people on which published data were to be based. The debate is likely not limited to France. On the whole, the French census is in good shape (chantal.madinier@insee.fr). Nevertheless, as a precaution and with a view to making it more effective, INSEE is developing new methods, presented here.

Survey on driving risk: The INRETS, the French National Institute of Research on Transport and Road Security is preparing the household longitudinal survey, MARC, on driving risk, driver mobility and behaviour. Sylvain Lassarre (<lassarre@inrets.fr>) in charge of this panel would very much appreciate getting inputs from IASS colleagues interested in this topic and in determining the most efficient longitudinal sampling in this kind of survey.

ITALY

by Claudio Quintano

The Italian National Statistical Institute (Istat) is carrying on a project to set up a **statistical register on Enterprise Groups**. A European Community regulation on Business Registers includes an option of information on Enterprise Groups and this may become compulsory in a few years time. Although participation in an enterprise group is a fundamental stratification variable for enterprises, the collection of statistical data presents many conceptual and methodological difficulties. Istat, in co-operation with other institutional bodies, has developed the conceptual framework of the register. Three administrative sources covering non-quoted companies, quoted companies, and consolidated accounts will be statistically integrated. The three sources provide information on relationships of shareholdings and relationships of control. The combination of different administrative sources will produce inconsistencies which need to be resolved before proceeding to the reconstruction of chains of control and enterprise groups. The chains of direct and indirect control will be constructed with a statistical algorithm developed by Istat. These chains will link the legal units belonging to groups to their nearest controlling unit and finally to the group head. As a result there will be two different outputs: a matrix containing data on control between each legal unit and its nearest controlling unit, and a matrix containing all the links between each legal unit and its group head. The matrix of nearest control is an essential tool which serves to keep track of the chain of control between the legal unit and its group head and to allow the drawing of different groups' perimeters for specific research purposes.

The treatment of foreign legal units belonging to transnational enterprise groups constitutes a specific and complex task. The statistical observation of transnational enterprise groups is essential for several reasons – they are trans-border actors since they control legal units situated in different countries, and, in a sense, they are the subjects organising external economic relations among the countries where they operate. The development of business registers on enterprise groups is thus an important target for the study of globalisation. This requires the ability to map transnational groups at the international level, but one of the main difficulties concerns the identification of foreign units within national business registers. Istat in co-operation with Eurostat is studying possible standards to be used by all European member states to record data on transnational enterprise groups in business

registers in order to obtain harmonised and comparable statistical data.

For more information, contact: *E. Morganti - Statistical Register of Non Agricultural Enterprises - ISTAT - Rome, Tel. + 39 06 72976223,* e-mail: enmorgan@istat.it.

According to a European Community regulation concerning the Structural Business Statistics, Member States provide Eurostat with provisional estimates (at 3 digit level of the Nace.rev.1 classification) no later than 10 months after the end of the reference year. This deadline does not allow Istat to complete its two main surveys of enterprises (sample survey of small and medium enterprises, and exhaustive survey of big enterprises). To respect the regulation and to produce reliable estimates, Istat decided to adopt a new production data strategy based on the **integrated use of administrative and statistical sources**. Such a choice is highlighted by the building of the Italian Business Register (ASIA), a high-quality and informative data base. In particular, the ability to access large administrative data bases containing economic variables (from the balance sheets data base managed by the Chambers of Commerce) and employment variables (from the social security agency data base) permits the re-engineering of the structural business statistics surveys. Balance sheets and employment micro-data of small and medium sized firms (less than 100 employees) have been integrated with data coming from the annual statistical survey on Provisional value-added estimates for large enterprises (over 100 employees). Therefore, the main issues to be developed are: i) implementing the integration of statistical and administrative sources, and ii) working on innovative methodological issues about the estimation techniques.

The estimation strategy adopted to produce preliminary estimates is based on a post stratification technique. The data base is a panel of enterprises and provides the economic and employment variables of interest. Specific procedures for improving the quality of the data have been applied. As a result, a 25,000 enterprises panel has been obtained. The data base used for the estimation of the variables of interest is not the product of a random sampling design. Therefore, post-stratified estimators have been used. The objectives of post-stratification techniques are: i) to provide data that represent subsets of the reference population; ii) to improve the estimator efficiency; iii) to reduce the bias resulting from the sampling design. Each stratum to which units are assigned is built according to auxiliary structural variables present in the Business Register and correlated to the characteristics of interest. In particular, it is assumed that an efficient estimate might be obtained through post-stratification by economic activity and by size. The post-stratification technique constrains the estimates to the number of employees included in the Business Register and allows the estimation of the variables requested.

That procedure has been used for the reference years 1997 and 1998. The number of employees provided by the 1997 ASIA Register (8,841,335) and the 1997 panel estimation results (8,899,205), have been compared. The model moderately overestimates the number of employees (+0.6%). The bias in the data base generates an overestimate of value added because of the characteristics of the selected enterprises. The assumption is that the same annual variation t/(t-1) of economic aggregates by sectorial breakdown provided by the panel data can also be applied to the final economic variables level from the traditional surveys. In this way, the final information concerning time t-1 can be updated at time t. This estimation procedure has been applied to 1997 sectorial data (NACE Rev. 1), resulting in forecasts of sectorial levels for the year 1998. By November 2000 a data base of administrative data including over 250,000 enterprises will be available. Its use will improve the precision of the estimates.

For more information, contact: *R. Monducci - Structural Statistics on Industrial and Services Enterprises - ISTAT - Rome, Tel. + 39 06 72976236, e-mail*: monducci@istat.it.

NEW ZEALAND

by Robert Templeton

Longitudinal Immigration Survey (LisNZ): The LisNZ is a longitudinal survey of immigrants to New Zealand that will be run in partnership between the Department of Labour and Statistics New Zealand (SNZ). The main aims of the survey are to assess the impact of Government immigration polices, and to describe the settlement experiences of migrants to New Zealand. The information gathered will be

used to identify factors which contribute to successful settlement processes. Of particular interest are the migrants employment, housing and health status, use of government agencies, integration into the community and financial status. Up to 7500 migrants will be selected and each will be interviewed three times: approximately 6 months, 18 months and 36 months after arrival in New Zealand, or after gaining residence if already on-shore. The first wave of the survey is planned to begin in 2002, with the full longitudinal analysis and final results of the whole survey being expected in 2008. For more information contact Judith Marbeck at 64-4-495-4600 ext: 4650, or email: judith_marbeck@stats.govt.nz.

Health of Maori Language Survey: SNZ will be running this post censal survey in 2001 for Te Puni Kokiri (Ministry of Maori Development). Development work is underway on this survey, which will sample people with Maori ethnicity as recorded in the 2001 Population Census. Development issues so far have focussed on the timing for sample selection, which is affected by population mobility after the Census and controlling overlap with other SNZ surveys. A pilot was run in May 2000, following the Census Dress rehearsal. This tested sample and interview techniques for the survey. For more information contact Paul Satherley at 64-4-495-4600 ext: 4716, or email: paul_satherley@stats.govt.nz.

Census of Businesses Involved in Biotechnology: The aim of this survey is to assess how much biotechnology activity is taking place in New Zealand and what future developments are planned. The survey will also collect economic and financial information on these businesses. The Ministry of Research in Science and Technology is funding the survey. The main development aspect was identifying the target population from SNZ's business frame and other sources. The survey was posted out in June 2000, and processing and data evaluation is now taking place. For more information contact Mark Jackson at 64-4-495-4600 ext: 8889, or email: mark_jackson@stats.govt.nz.

Business Practice Survey: This survey is funded by the Ministry of Economic Development and Ministry of Research, Science and Technology and aims to gain an understanding of the wider state of business and management practice in NZ firms. This evolved from an initial proposal to survey businesses about use of information technology, in particular e-commerce. The survey is expected to be in the field in February 2001. For more information contact Mark Jackson at 64-4-495-4600 ext: 8889, or email: mark_jackson@stats.govt.nz.

Data Recognition Testing for Population Census: Data recognition will be carried out in the 2001 Census of Population and Dwellings. Numeric and tickbox recognition were used in the 1996 Census. These will be used again in the 2001 Census and it is likely alphabetic recognition will be used for the first time. The Census dress rehearsal in March 2000 contained 3 alpha recognition questions - occupation, language and religion, and evaluation is now taking place. Since this was mentioned in the Country Reports in June 1999, extensive testing has taken place on alpha-numeric recognition. For more information contact Tracy Moore at 64-4-495-4600 ext: 8757, or email: tracy_moore@stats.govt.nz.

PHILIPPINES

by Gervacio G. Selda, Jr.

The Institute for Labor Studies (ILS) and the Bureau of Labor and Employment Statistics (BLES) of the Department of Labor and Employment (DOLE) conducted the **Philippine Labor Flexibility Survey 2000** last August 2000. The survey is intended to update the data collected in 1990 for use in coming up with more responsive labor-related policy and program in the light of the increasing prevalence of flexible forms of employment in the Philippines brought about by rapidly globalizing economy. Where flexible forms of employment are getting to be the "norm" rather than the "exception", the continuing relevance of labor laws and the whole gamut of labor institutions which have been largely hinged on the standard employer-employee relationship may have to be seriously reassessed, to ensure the welfare and protection of workers under such forms. Also, in view of the growing commitment toward gender equality in the workplace, the gender dimension of the flexibilization and other forms of discrimination would have to be dealt with more extensively. Specifically, the study aims to generate updated information on (1) establishment profiles; (2) nature and condition of flexible work; (3) gender-sensitivity in establishments; and (4) application and relevance of labor-related laws. Data collection is through mail/self-completion questionnaire and

personal interviews of 1,366 establishments taken from the 1998 National Statistics Office (NSO) updated list of establishments, stratified according to their location, economic activity and size of employment and will be completed this October 2000. Results of the survey are expected to be available by January 2001. (For more details, contact: Ms. Merle C. Cabaero, Chief of Worker's Welfare Research Division, ILS, DOLE, Intramuros, Manila, Philippines, Fax No. (632) 527-3448, e-mail: ilsdole@ info.com.ph).

The Department of Transportation and Communications (DOTC) has conducted the first part of the **Air Passenger Systems Survey** last August 2000. The survey is designed to: (1) develop and refine a workable system of gathering information of air passengers in some selected airports in the Philippines which can be replicated once it is put in place; (2) update the 1990 origin-destination of air passengers whose commercial air travel commenced at various points; and (3) develop a computerized database on the origins and destinations of air passengers whose commercial air travel commenced at various airports. Since different airports handle different levels of passenger volume, the probability proportionate to size (PPS) sampling plan will be followed to account for these differences and provide for a more accurate profile of the entire passenger population. The total sample size of the survey is 6,000 with 2,400 to be interviewed during the lean season (August) and 3,600 to be interviewed during the peak season (December). The main types of information to be gathered for the study include the following: (1) true origin and destination of the passenger; (2) purpose of the trip; (3) frequency of travel; (4) means of travel to the airport; (5) demographic profiles of the respondents; (6) perceptions regarding the airport facilities; (7) perceptions regarding the airline facilities; (8) perceptions regarding the airline industry; (9) number of companions accompanying the traveler to the airport; (10) number of persons traveling with the respondent; and (11) suggestions for airport facility improvement. The second part of the survey will be conducted in December 2000. The resulting data will be analyzed using SAS/STAT software and subjected to various types of multivariate data analysis to determine the relationships between the different variables that are generated by the survey. (For more details, contact: Ms. Malou Gicana, Information System Department, DOTC, Ortigas Avenue, Mandaluyong City, Philippines, Fax No. (632) 725-0025).

The UNICEF Philippines in coordination with the Country Program for Children (CPC) V Provincial and City Governments of the Philippines will conduct in October the **2000 CPC V Provinces and Cities Multiple Indicator Cluster Survey** (MICS). The 2000 MICS is a follow-up of the 1999 MICS. Unlike the 1999 MICS where the scope is national, 2000 MICS will cover only CPC V provinces and cities. The main objective of the survey is to establish benchmark information on these CPC areas. Likewise, the survey is intended to validate values obtained in the 1999 MICS for use in the End Decade Assessment on the achievement of goals for women and children. The MICS 2000 will use the Expanded Program of Immunization (EPI) cluster sampling design with the provinces and selected cities as domain. The design is 2-stage with barangay (smallest political unit in the Philippines) as primary sampling unit and cluster of households as secondary and ultimate sampling units. The barangays will be selected with systematic probability proportional to size sampling. The size measure is the number of households in the barangay. The barangays are listed according to rural-urban classifications and/or ecosystem(s) present (coastal, upland, islands, etc.) to induce implicit stratification. The sample barangay will be divided into areas/segments of household clusters of approximately similar size (usually would range from 15 to 45 households per cluster). Then one sample is selected with equal probability sampling. All households in the sampled cluster are enumerated. The major data items to be collected are: Vitamin A coverage, Iodized salt consumption, ECD, school enrolment rate, school attendance, educational attainment, child labor, maternal and infant health, care of illness (diarrhea, etc.), knowledge of HIV/AIDS, socio-economic profile, birth registration and early learning, and child rights. (For details, contact: Mr. Augusto Rodriguez, Information Management Officer, UNICEF, NEDA sa Makati Building, Makati City, Philippines, Fax No. (632) 892-6456, e-mail: arodriguez@unicef.org).

POLAND

by Janusz Wywial

In Poznan in 1994 the Centre of Regional Statistics (http://csr.ae.poznan.pl) was established by the Central Statistical Office in Warsaw and the Poznan

University of Economics. About 30 scientific researchers from the following institutions: the Poznan University of Economics, the Poznan University of Agriculture, the Adam Mickiewicz University, the Statistical Office in Poznan, and regional and local administration offices, work in the Centre of Regional Statistics. The Centre is managed by Jan Paradysz e-mail: paradysz@novci1.ae.poznan.pl, paradysz@csr.ae.poznan.pl.

Within six years of its creation the Centre has worked out and is still improving the methodology of the regional survey. The research experiences of the Centre include the methodology of indirect estimation (small area statistics), and data processing in geographical information systems and database systems.

The main trend of research in the Centre is connected with Polish information infrastructure. Studies in this field include:

a. evaluation of the existing sources of information supply at the lowest possible level of territorial aggregation such as censuses, agricultural censuses, Labour Force Survey, computer databases on labour market and firms, population registers, the local databank,
b. searching for new, unconventional information sources for public statistics such as the geodesic database for numerical maps (Geo-Info), real-estate cadastre, the tax information database (POLTAX), the vehicles register and others,
c. integration of computer databases,
d. usage of Geographic Information Systems in the presentation of regional data,
e. integration of computer databases in the example of Geo-Info,
f. usage of enterprises register (REGON and DG1) as a source of information about the economic activity of medium-sized enterprises,
g. small area statistics,
h. regional agricultural statistics,
i. cadastre as the data source for regional statistics,
j. demographic processes at the micro-level,
k. Internet Panorama of Communes, Cities and Regions – the Small Fatherlands promotion.

The scientific projects carried out by the Centre, and by the people connected with the Centre, are presented through meetings. The participants at these meetings are not only people connected with the Centre in the scientific sense but also people from outside who are interested in the subjects covered. The Poznan Centre of Regional Statistics was the organiser of four conferences on regional statistics in 1994, 1996, 1998 and 2000. The results of the first three conferences were published in seminar proceedings.

The conference on Regional Statistics for local government and business was held June 5-7, 2000 in Poznan- Kiekrz. The following topics were covered: theory and practice of survey methodology, small area statistics, labour market, classification and data analysis in regional cross-section, new and potential data sources for regional statistics, Internet Panorama of Cities, Communes and Regions - Small Fatherlands promotion.

The Conference on the Application of Small Area Statistics in Official Statistics was held on March 28-29, 2000 in Radom. Several proposals for the estimation of economic characteristics (including labour force) of small regions by means of small area sampling techniques were considered. For more details, contact J. Kordos at J.Kordos@stat.gov.pl and E. Golata at golata@novci1.ae.poznan.pl.

SOUTH AFRICA

by Jairo Arrow

In September Statistics SA released a new publication, **Measuring Poverty in South Africa**. The publication contains four papers which draw on the agency's data to analyse poverty and inequality patterns in the country. The papers are preceded by a commentary by Professor Charles Simkins of the University of Witwatersrand. The commentary reviews the methodological approaches of the papers and points out the strengths and limitations of each.

The first paper combines census and survey data to construct a poverty map of the country. The analysts drew on data from the five-yearly income and expenditure survey and the annual household survey, and used regression analysis to impute expenditure values for each household in the census. One of the outputs is a map, included in a pocket in the back of the publication, showing

poverty levels across the country at the sub-provincial level.

The second paper describes the construction of two indices – a household infrastructure index and a household circumstances index. The indices are constructed on the basis of census data. The first index gives an indication of the meeting of basic needs. The second is related to empowerment. The methodology used to isolate the two indices includes a principal components factor analysis.

The third paper focuses on inequality in earned monetary income. The paper uses data from four consecutive annual household surveys to compute gini coefficients for the population as a whole, as well as for each of the four population groups defined by the apartheid regime. The analysis suggests some increase in inequality over the four years. The paper also discusses gender differences in the experience and incidence of poverty.

The final paper describes the calculation of a social accounting matrix (SAM) based on household income. This chapter draws on the economic statistics used in the calculation of national accounts, but focuses on households rather than institutions. Statistics SA is currently constructing a new SAM, based on the 1993 modifications to the System of National Accounts. The new model should be released in 2003.

For further information, contact Jairo Arrow, Statistics South Africa (jairoa@statssa.pwv.gov.za).

SPAIN

by Rosa Maria Bermudez Gomes

Together with its new *Welcome,* the **INE launches a statistical information storage system on the Internet**, with the name of INEBASE. INEBASE is loaded with all the information that is produced in electronic formats.

The **2001-2004 National Statistical Plan** has been approved. It outlines all the statistics that the General Administration of the State aims to carry out in the forthcoming four years with an indication of the budget for each of them.

A study has been issued on the ***Use of New Technologies in the Manufacturing Sector***. It is the first complete research carried out in Spain on how top technology is applied in the manufacturing enterprises and on its short and medium term results.

In the **environmental sector**, the INE presented last July the advance results of the survey on water supply and treatment as well as of the 1998 survey on the collection of urban solid waste. At present, investigations are underway on the generation of dangerous and not-dangerous waste and on the use of water in the agricultural sector.

The objective of the **Hotel Price Index** (HPI) is to know the seasonal evolution of prices charged in the hotel business to any type of customers. These may be not only households but also enterprises (whose staff consumes accommodation services for its business travels), tour operators and travel agencies (which are not the final consumers of the accommodation services and may be considered intermediaries of the tourist sector). In this price index, the structure of the tariff weighting remains fixed.

Complementary to the Hotel Price Index, the **Hotel Income Index** has been elaborated. In analogy with the former, this index may be defined as the relation between income from accommodation services for a double room with bath during the target period and the base period, taking into account the varying application percentages of all the tariffs in the different periods. It is, therefore, a value index which compares the income in two periods in which the prices as well as the weight structure of the tariffs vary.

Since March 2000, the INE publishes quarterly information on the evolution of **Electronic Trade** in such an important sector as retail trade. Information is provided not only on the geographical distribution of electronic trade but also on the goods commercialized by the enterprise, broken down by food, personal equipment, household equipment, health, leisure and other goods. Information is also available on the different objectives of electronic trade: advertising, relationship with the customers, sales, purchases and other issues.

For any further information please contact Rosa Mª Bermúdez, Instituto Nacional de Estadística, Paseo de la Castellana 183, 28046 Madrid, Spain. E-mail: rbermudez@ine.es.

SPAIN – BASQUE COUNTRY

by Martin Gonzalez Hernandez

Eustat, the Basque Statistics Institute, is developing a project under the title The Population Register, through which different sources of administrative and statistical information are integrated. This serves as the basis for statistics produced by the Institute, above all, population statistics. At present the main sources of administrative information are the Municipal Registers. Birth, death and marriage data are obtained from the Civil Registers and divorce figures come from the courts. The main source of statistical information comes from Population and Housing Censuses, which are the only sources from which data are obtained on the socio-economic characteristics of people and the structure of housing. Though it is also possible to integrate information from other origins, mainly administrative, initially the Population Register will be developed using the above sources.

The Population Register is structured through two Modules of related data:

1. The Province database Module in which statistical units are located in terms of space as companies, housing and the population residing there. Through this Module data can be obtained at concrete and very diverse territorial levels such as municipality, section, neighbourhood, building, etc.; and
2. The Housing and Population database Module, which to respond to the existence of two different units, is divided in two modules: housing and population.

Information from different sources will be integrated into these Modules, but referring to the units themselves: population, housing and province. Integration means that current statistics can contain more information than the statistics from which their results were obtained and, consequently, represents an enhancement of their possibilities. In addition, the updating that integration of sources with periodic information involves also redounds in the quality of results. This includes Migratory Movements Statistics based on 1999 data and Births, Marriages and Deaths Statistics from the reference year 2000. Divorce statistics are also produced. Base data on divorces had been available previously, but were impossible to develop until integrated with data from the Civil Registers on marriages.

Another great advantage lies in the possibility of obtaining new statistics in which the time dimension, i.e., longitudinal analyses, is fundamental. But in turn, the integration of new information sources, whether Administrative Registers or not, makes the possibilities available through an integrated model of information immense. Thus the most important facts in the life of a person or a family may be reconstructed and statistics may be developed with a longitudinal or diachronic vision, without having to carry out expensive investment in new field operations.

For sample surveys, the improvement will come from a more up-to-date knowledge of the universe and a radical improvement in the identification of the surveyed units and their relationship with this universe. Also, it can help to reduce the number of questions put to interviewees on, for example, demographic topics. Another use can consist of the determination (or imputation) of the characteristics of people who do not respond in the surveys, for example, on occupation, activity branch, level of knowledge of the Autonomous Community language, etc.

For further information, contact Martín González Hernández, Head of Demographic Statistics, tel: 34 45 01 75 49, E-mail: Martin_Gonzalez@eustat.es.

Use of Business Income Tax Data to Extend the Information Available from the ABS Economy Wide Economic Activity Survey

Paul Sutcliffe & Steve Crabb, Australian Bureau of Statistics
Paul Sutcliffe, Australian Bureau of Statistics, P.O. Box 10 Belconnen ACT 2616, Australia
p.sutcliffe@abs.gov.au

The views expressed in this paper are those of the authors and do not necessarily reflect those of the Australian Bureau of Statistics (ABS).

1. Introduction

In the last decade the Australian Bureau of Statistics (ABS), like many other official statistical agencies, has been under increasing pressure to increase the range of statistics it makes available; maintain and improve the quality of these statistics; and provide them in timely fashion, while becoming more cost effective. In particular, in the last few years the need to reduce the statistical reporting burden placed on businesses has led to more effective uses of administrative data, especially business income tax data provided by the Australian Taxation Office (ATO). The Economic Activity Survey (EAS), which commenced in the early 1990s, has recently been modified to use taxation data to supplement information collected directly from businesses.

2. Methodology

Currently EAS is designed to provide two main types of data output. Firstly, a single-phase survey collects financial data (base data) such as main income and expenditure items, which are readily available from the annual accounts of each business. Secondly, a smaller second-phase subsample collects detailed income and expenditure data required for the Input-Output (I-O) Survey, which is used in the compilation of the national accounts. Traditionally these data have been collected directly from businesses.

Incorporation of taxation data into the design for the service industries component of EAS is achieved through the partitioning of the economy wide population framework into subpopulations from which either direct collection or administrative data, or both, can be used. The larger and more complex businesses are only surveyed using direct collection. However, for some classes of businesses, the directly collected EAS data (used in the preliminary publication) and the taxation data are closely comparable at the unit level.

In order to effectively operationalise the use of the two sources of data, the population has been divided into four streams:

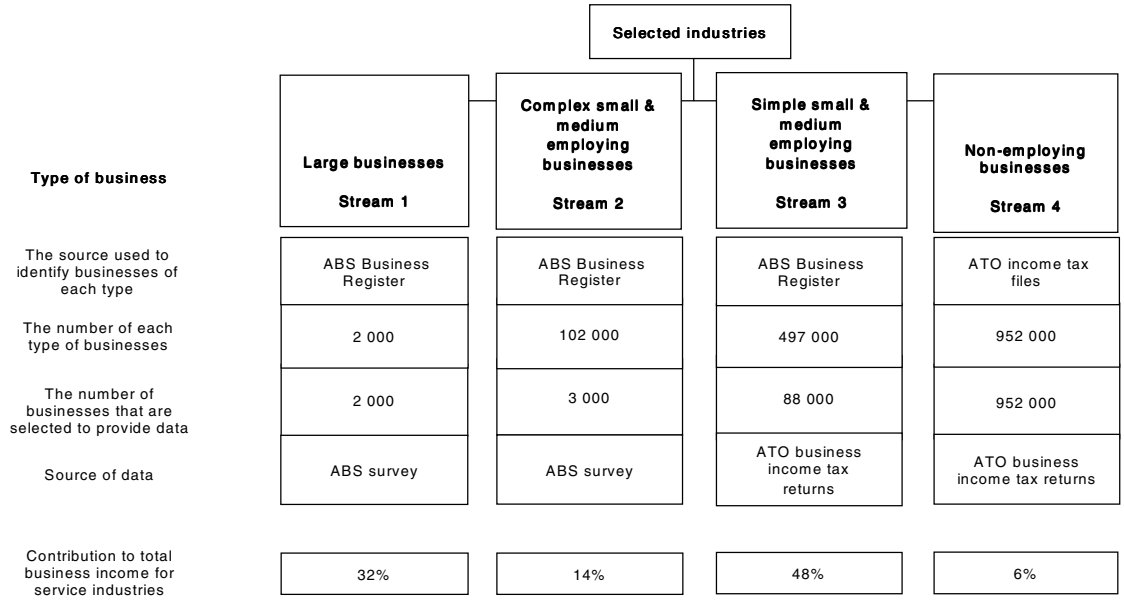
- large businesses that have complex structure (stream 1);
- complex small- and medium-employing businesses (stream 2);
- simple small and medium businesses suitable for income tax data substitution (stream 3); and
- non-employing businesses, which are only available from the income tax files (stream 4).

Table 1 shows the relationship of the four streams.

For the final EAS publication a large sample (100,000) is selected from the ABS Business Register. The selected sample units are then matched against taxation records. Because the data items available from tax are limited, a much smaller "second-phase" sample is directly approached so that the full set of base financial data can be published. The current first-phase sample is designed to achieve smaller relative standard errors (3%) at the ANZSIC subdivision level than the existing and directly collected EAS (7%), which is used for the preliminary publication.

The use of tax data is not without its problems. In our case these surface when estimating the benchmark total for the first phase $\hat{X}_{Tax}$.

Table 1. Summary of data sources for the EAS/tax strategy



Once the first-phase sample has been selected from the ABS Business Register, the strategy then requires the selected sample to be matched against the appropriate taxation file. In an effort to produce timely statistics, the ABS uses information from business income tax files that are generated 12 months after the end of the financial year. At this point in time not all businesses would have provided their income tax returns. Currently, indications are that income tax returns for approximately 10% of businesses would still be outstanding. Additionally, a proportion of businesses that are included on the ABS Business Register would not have traded for the year in question. After taking these factors into account, it is estimated that data for between 15% and 20% of live businesses selected in the sample are not available for processing when estimates are produced. Data for these unmatched businesses must, therefore, be estimated.

In practice, studies have shown that the first-phase selections, which match to the taxation data, have a higher proportion of live businesses than the nonmatched units. Thus, the nonmatched units cannot be considered to be "standard nonrespondent" units and their values should not be estimated from those of the respondent units. To account for this "response bias" the estimates are adjusted for the lower proportions of nonmatched units that are live using an adjustment factor. This then simplifies to an adjusted simple expansion estimator:

$$\hat{X}_{Tax} = \sum_{h=1}^{H} \frac{N_h}{n_h}\left[1 + f_{sm}\frac{n_{h\bar{m}}}{n_{hm}}\right]\sum_{i=1}^{m} X_{hi} ,$$

where $f_{sm} = \frac{n_{h\bar{m}}}{n_{hm}}$ is known as the adjustment factor and *s* represents the level at which the factor is calculated. Note that, when there is full matching between the selected sample from the Business Register and the taxation records, this factor is zero and the estimator reduces to the standard simple expansion estimator.

The first-phase sample provides accurate benchmark data while the second-phase sample collects detailed breakdowns that can be used to prorate the benchmarks for output purposes. The prorating factors refer to the different benchmark variables, which are used for each type of output.

For example, in stream 3 (simple small- and medium-employing businesses) the first-phase estimates of total income are calculated using the data from taxation records, but estimates of all dissections, or components of total income, are estimated using prorating factors. Total income is also collected in EAS, and the factors estimated

from EAS are the proportion each component contributes to the total. Thus the prorated estimates are

$$\hat{y}_{pro\_rated} = \frac{\hat{y}_{EAS}}{\hat{x}_{EAS}} \hat{x}_{Tax},$$

which is the standard form of the two-phase ratio estimator.

While the estimator outlined above is in the standard form of a two-phase ratio estimator, it should be noted that the benchmark information collected from tax records in the first phase and that from EAS in the second phase come from different sources. Therefore, at the businesses level, reported values may not be equal for the same base data item and so the population totals are not necessarily equal ($X_{Tax} \neq X_{EAS}$) as would be the case in a traditional two-phase survey.

The variance estimator for the two-phase estimator requires modification to account for the variance associated with the matching of taxation data in the first phase. However, the current variance estimation procedures for the first phase ignore this random nature of the adjustment factor and treat f_{sm} as a constant.

A recent study showed that the estimated variance for stream 3 increased when the random nature of the adjustment factor, f_{sm}, is taken into account. This resulted in the overall relative standard errors increasing from 3% to 3.9%. These are significant changes in the RSE's. One must thus consider the application of the correct formula in variance estimation procedures. Therefore, we are currently investigating the use of the jackknife variance estimation technique used by the ABS generalized estimation system (**GENEST**).

3. Challenges for the Future

Supplementation of the Economic Activity Survey with business income tax data has enabled the ABS to improve the range and quality of statistical output produced while at the same time minimizing the reporting load on providers. The main challenges for the future are to utilize tax data in satisfying other areas of unmet demand such as finer industry statistics and geographic breakdowns and to promote the use of business income tax data by other collection areas in the ABS.

While the methodology described in this paper has been successful in improving the quality of industry data available regarding small- and medium-sized businesses, user demand remains for data at finer levels of industry classification, and in particular 4-digit ANZSIC Class level data. The current methodology employed limits the availability of finer level industry output due to the sample design in the small- and medium-employing business streams (Streams 2 and 3 in Table 1). Future work proposed by the ABS will target those industries for which reliable Class level results could be obtained using tax data. For instance, for those industries dominated by simply structured small- and medium-employing business or non-employing businesses, an increase in the tax sample of small and medium business may result in reliable Class level estimates.

A significant influence on the future use of business income tax data in the ABS will be the current process of tax reform in Australia. Tax reform covers many changes to the tax system including the introduction of an Australian Business Register (ABR), the introduction of a Goods and Services Tax, a review of business taxation and introduction of a new Pay as You Go system for the payment of income and other taxes. The reform process provides many statistical opportunities, including the availability of regular Business Activity Statements (BAS). The BAS is a new government requirement for businesses to claim credits from the new tax system. The statements contain information about turnover, wages paid, capital expenditure, and exports and other expenses and is available on a quarterly basis. The BAS, while not containing the detail of a business income tax form, will be available on a far more timely basis and should provide a rich source of information with which to supplement ABS direct collection surveys in the future.

Statistics New Zealand's Experiences in the Use of Taxation Data for Industry Statistics

Anne McAllister, Senior Economic Statistician
Business Respondents and Data Sources Management
Statistics New Zealand

This paper is a summary of the one presented to the 13th Conference of Commonwealth Statisticians, Botswana, 2000. It discusses the progress that Statistics New Zealand has made in using tax data from the tax agency, Inland Revenue Department (IRD); it outlines future developments and discusses some of the problems and risks that we have encountered.

Examples of the use of tax data by Statistics New Zealand follow:

1. NZ Business Register Maintenance

Tax data are used to identify and maintain information about businesses for business register purposes. New businesses are surveyed to collect additional information such as employment for stratification and estimation and imputation purposes and to confirm IRD information about contact details, business activity descriptions, etc. The current maintenance strategy has been reviewed. It is proposed to make extensive use of business activity descriptions and tax returns data such as Goods and Services Tax (**GST)** returns and monthly employer deductions data for size indicator purposes.

2. Statistical Outputs

GST returns have been used to produce a monthly Business Activity Indicator **(BAI)** series. It has provided monthly sales and purchases by industry since April 1992. A regional series is currently being developed using employment information from the business register to apportion enterprise data to regions. There are a number of weaknesses in the series. These include timeliness delays, group returns that do not match the business register statistical units, differing return filing frequency depending on the size of the business, and inclusion of capital sales and purchases transactions. These problems have resulted in the data being released as an experimental series rather than an official series.

These deficiencies are partly being addressed in the current sub-annual survey development. The retail and wholesale trade and manufacturing surveys are currently being redesigned, and coverage will be extended to other industries to produce quarterly economy-wide economic statistics on nominal income, expenses, inventories, and capital formation. It is proposed to use **GST** as a stratification variable for large-sized enterprises and use temporally apportioned data from the **BAI** for small- and medium-sized enterprises. Some estimation will be needed for the last month in the quarter as **GST** data will not be available in time to use in the release.

Data from annual financial accounts summary returns were used successfully in the 1995 Manufacturing Census and the 1996 Rest of Economy Census to reduce the number of units requiring to be directly surveyed. Lessons learned in using tax data in the censuses were applied to the 1998 Annual Enterprise Survey redesign. The annual accounts returns were successfully used to provide estimates for all un-incorporated businesses. They were also used in their entirety to provide data for the agriculture and commercial property industries. It is hoped to extend their use to other small- and medium-sized enterprises as development funding permits.

Monthly employer deduction returns have been matched with the business register to produce a test database of monthly gross earnings by industry and region. The data are currently being analysed to determine whether they are of acceptable quality to be published as survey output in their own right.

These returns are likely to be used to provide a wages and salaries estimate for small- and medium-sized businesses in the sub-annual survey development.

Risks in relying on a third party for data

As shown in the earlier examples, Statistics New Zealand has had considerable success with tax-sourced data in terms of business register maintenance and as a substitute to direct collection in surveys. It is seen as playing a key role in our economic data collection strategy. Significant reduction in the number of businesses being directly surveyed has already been achieved.

However, some of the potential advances in the exploitation of tax-sourced data will require some changes to be made by the tax agency in order for these to be realised. The application of tax data is often limited because of a poor fit to statistical needs. Examples include an incompatibility between the tax reporting unit, which the return represents, and the statistical unit. This applies particularly to group enterprises and impacts on the quality of industry and area estimates. Also, with respect to sub-annual data there can be differences between the reporting frequencies as well as timeliness issues.

Managing the relationship between the tax and statistical agencies is very important. This is essential to ensure both continuity and quality of supply. Fostering close cooperation in order to arrange for changes to be made by the tax agency so that tax-sourced data will better meet statistical needs has the support of government. It encourages such initiatives that will contribute to a reduction in the compliance costs incurred by businesses.

There will continue to be a risk in using tax data especially with Inland Revenue's drive to simplify tax collection and reduce compliance costs on its customers. However, moves by Inland Revenue to automate processes and encourage taxpayers to submit electronic returns provide some new opportunities for Statistics New Zealand. Greater penetration of user-friendly financial accounting and payroll packages may allow us to collect information directly from respondents without significantly affecting their workload. This data are more likely to be a better fit with Statistics New Zealand data requirements and will offer significant improvement in timeliness.

Special Articles: Censuses Conducted around the World

ALGERIA: SAMPLE FRAME OF THE 1998 CENSUS OF POPULATION AND HOUSING

Algeria is divided into 1,541 communes (administrative distribution of 1985). For statistical purposes, each commune is composed of agglomerations and/or a sparsely populated area. An agglomeration is a group of buildings (minimum 100) separated by less than 200 metres. We distinguish between the seat of the communal popular assembly (CPA), which we call the "agglomération chef-lieu" (principal agglomeration) (1,487), and the other agglomerations, known as "agglomérations secondaires" (secondary agglomerations) (2,563). Hence there are 54 communes with no principal agglomeration; instead, they have a principal hamlet (a settlement with fewer than 100 buildings).

These agglomerations (a total of 4,056, 16 of which are empty) and the sparsely populated areas of communes are divided into enumeration areas or districts (41,021).

There is a sparsely populated area in 1,492 of the 1,541 communes (16 of which are empty; that is, there are buildings, but there were no inhabitants on the day before June 25, 1998).

Conurbation is taking place; that is, the agglomerations of two or more communes are merging. In particular, major urban centres are merging with their suburbs. There are 13 such centres containing 55 agglomerations (28 in Algiers alone).

A number of criteria are used to distinguish between urban and rural. In Algeria, the distinction is purely statistical, whereas in other countries, it has an administrative basis. Consequently, Algeria does not differentiate between urban and rural communes but rather between urban and rural agglomerations with a number of possible variants. The sparsely populated area is automatically classified as rural.

To qualify as urban, an area must meet the following five criteria:

- A **minimum population** of **5,000**: A truly urban way of life is unlikely below this threshold.
- Economic activity: In our view, the fundamental characteristic of a city is that at most only a small proportion of its total working population (**less than 25%**) is engaged in **agriculture**. The preliminary results of the 1998 census provide a breakdown of the employed population by sector (agriculture, other sectors).
- Urban characteristics: Some features are considered to be exclusively urban in nature: public services, hospitals or clinics, educational institutions, law courts, and entertainment centres (movie houses, theatres, etc.).
- Connections to public utilities (drinking water, sewerage, electricity). In some cases, we used data from the 1995 communal survey (in addition to census data) to determine whether specific agglomerations had the required facilities. In addition to the minimum population (5,000) and the 75% non-agricultural requirement, the agglomerations had to meet the following conditions:
  - connection to the **drinking water system**;
  - connection to the **electric power grid**;
  - connection to the **sewerage system,** and
  - at least **three** of the following **five** requirements:
    . a hospital or clinic;
    . a secondary school or college;
    . socio-cultural facilities (daycare centres, youth centres, other facilities);
    . a sports and recreation infrastructure (stadiums, amusement parks, movie houses, theatres);
    . government services (post offices, law courts, government offices).
- The average annual rate of population increase between 1987 and 1998 provided information about the agglomeration's ability to attract migrants. This characteristic was used with discretion. We considered that a high rate of increase was not necessarily a good indicator that an agglomeration should be considered urban.

Finally, for the aforementioned reasons, administrative rank was not considered as a criterion, as it was in 1977 and 1987.

On the basis of the above variables, the 579 agglomerations were assigned to four different strata:

1. **Large urban:** This stratum contains 24 agglomerations, each of which has **over 10,000 people employed, more than 75%** of them in **non-agricultural** industries. These agglomerations have a large number of high-level services: higher education, specialized hospitals, a high coverage rate for basic infrastructure.
2. **Urban:** This stratum contains 120 agglomerations, each with a population of **at least 20,000**. **More than 2,000 people** must be employed in **non-agricultural** industries, and they must account for at least 75% of total employment. Unlike large urban agglomerations, they have only a small number of high-level services.
3. **Suburban:** The 97 agglomerations in this stratum are neighbouring residential areas. They are extensions of the residential areas – and in some cases the industrial areas – of the four metropolitan areas. In general, they are very close to the **four regional metropolitan areas** (**satellite cities**). They meet the industry prerequisite, and they have the primary urban characteristics.
4. **Semi-urban:** This stratum contains 338 agglomerations (one more is currently under construction). They must have **at least 1,000 people employed in non-agricultural industries** and a **minimum population of 5,000**. This stratum is composed of units that influence the surrounding area and provide minimum services, especially in education and health, not only to their own populations but also to nearby rural units (semi-rural areas, rural agglomerations and sparsely populated areas).

Distribution of agglomerations by type

	1	%	2
Large urban	72	1.3	24
Urban	127	2.3	120
Suburban	97	1.7	97
Semi-urban	339	6.1	339
Urban	**635**	**11.4**	**580**
Semi-rural	480	8.7	
Rural agglomeration	2941	53.0	
Rural	**4913**	**88.6**	
Sparsely populated area	1492	26.9	
Total	**5548**	**100.0**	

(1) Excluding intercommunal agglomerations.
(2) Including intercommunal agglomerations.

Since there is no direct means of determining the types of communes, we used an automatic classification system based on territorial composition.

Distribution of communes by type

Type of commune	No.
Large urban	59
Urban	127
Suburban	77
Semi-urban	322
URBAN	**585**
Semi-rural	130
Rural agglomeration	382
Sparsely populated area	444
RURAL	**956**
Total	**1541**

This table shows predominant characteristics. That is, we have 59 communes that are predominantly of the "large urban" type, 127 that are predominantly urban, 77 that are predominantly suburban, and so on.

In fact, we have the following:

- 31 completely urban communes (no sparsely populated area);
- 954 completely rural communes, of which:
  - 13 consist solely of rural agglomerations with no sparsely populated area;
  - 902 have one or more rural agglomerations and a sparsely populated area;
  - 39 consist entirely of sparsely populated areas;
- 566 mixed communes, of which:
  - 516 have one urban agglomeration and a sparsely populated area;
  - 35 have more than one urban agglomeration and a sparsely populated area.

CANADA'S 2001 CENSUS OF POPULATION

In Canada, a Census of Population is conducted every five years. Canada's first Census of the third millennium will take place on May 15, 2001. This short description will give an overview of the Census and will outline major changes and improvements that have been implemented since the last Census in 1996 as well as plans for the 2006 Census.

The efficacy of traditional census methodologies in a changing environment is increasingly debated by many of its constituencies. Expectations of the Census and its products have changed and grown. Concurrently, in an environment of continued fiscal constraint, the funding agencies are pressing to look at various alternatives to the Census and efficiencies within the Census.

Starting in 1995 with a Strategic Planning Conference and using the evaluation of the 1996 Census - which included tests of potential methods for 2001 - as well as various stages of testing, Canada's Census fundamentals have been questioned and aspects such as roles, values, methods, technologies and products were assessed against available options. Since then, wherever feasible, changes have been implemented to improve the cost-effectiveness and to ensure the relevance, accuracy, timeliness and accessibility of the Census and its results. In this context, Statistics Canada has continued for 2001 with a process of incremental change. More importantly however, these changes are being made as part of an overall set of strategic objectives which the Census has set for itself.

The Census covers the entire Canadian population, which consists of Canadian citizens, landed immigrants and non-permanent residents (persons living in Canada who have a student or employment authorization, a Minister's permit or who are claiming refugee status), together with family members living with them in Canada as well as Canadians living abroad in military camps or working at embassies.

The 2001 Canadian Census will primarily use a dropoff/mailback collection methodology to enumerate a population of about 31,000,000 in about 12,000,000 households using about 32,000 enumerators. Mailback is to the enumerators for checking and follow-up. As in the past, basic demographic data will be collected on a 100% basis while a more detailed questionnaire will be given to a one in five sample of dwellings. In a large scale test, the option to respond via the Internet will be offered to 200,000 dwellings in two test sites. For the third time Statistics Canada's Address Register will be used by enumerators as a coverage improvement tool. Data capture (via key entry) will again take place at a number of regional processing sites with all subsequent processing being done in Ottawa. In a new development, images of all questionnaires and enumeration area visitation records will be captured and used in subsequent steps where manual processing is required, as well as for archiving purposes. Most coding of variables will be automated with computer assisted resolution of problem cases using the questionnaire images. The Nearest-Neighbour Imputation Methodology (NIM), first used in the 1996 Census for a number of demographic variables, will be used in 2001 for an expanded set of variables. A calibration methodology will again be used to weight the sample to 100% data totals.

As in the last several Canadian censuses, coverage will be measured using the reverse record check methodology in which persons in the target population identified through the last census, birth records, and immigration records are traced to determine whether they were enumerated in the current Census. Census dissemination plans call for the public release of population counts in April 2002, with the release of other variables occurring between July 2002 and May 2003. Several variables are to be released earlier in the cycle than for the 1996 Census, with the final release occurring one month earlier. Population estimates adjusted for net Census undercoverage will be released in September 2003. With the 1996 Census, Census data made its first presence on the Internet and print publications evolved towards CD-ROM publication. This direction will be pursued further for 2001.

Canada's Census of Agriculture is conducted in conjunction with the Census of Population. At the point of dropoff of the Census of Population questionnaire, Census of Agriculture questionnaires are also distributed at dwellings identified as including a farm operator. They are also distributed to all additional households including a farm operator when indicated on the Census of Population questionnaire.

The remainder of this article focusses on the Census of Population and develops the subjects noted in the previous paragraphs in more detail. They will be presented approximately in chronological order as they occur in the Census.

1. Content Determination

Before each Census, extensive consultation is conducted with decision makers, analysts and data users for their views on the type and extent of information that should be available. Building on the consultation experience of the 1991 and 1996 Censuses, a consultation guide was developed to help Canadians contribute their ideas and suggestions about the content of the 2001 Census questionnaire. A summary of the process, findings and follow-up activities can be found in the *2001 Census Consultation Report* which is available on Statistics Canada's Internet site at www.statcan.ca.

An extensive content testing program was undertaken. A national pre-test of four versions of the short questionnaire was conducted in May 1997. As well, before new questions or modifications to previously asked questions are used, they undergo extensive qualitative (cognitive) testing. In many instances wording and instructions on the Census questionnaire were modified based on these results.

An important part of the preparation for the 2001 Census was the 1998 National Census Test (NCT) conducted in 13 sites across Canada, covering about 77,000 households. Traditionally, the NCT is used to test potential new census questions as well as possible changes to existing questions. For the 1998 NCT a new collection methodology involving mailback to a centralized location for edit and follow-up, first tested in one region of Canada in the 1996 Census, was also tested.

In the 1998 NCT, one third of the households sampled were given a short form, one third received a first version of the long form and the other third received a second version of the long form. Compared with the 1996 Census, the 1998 NCT included six additional questions on birthplace of parents and of grandparents, secondary languages spoken at home, language at work, language of education and place of study. Two decennial questions on fertility and religion, last asked in the 1991 Census, were included. The test also included variations of questions on ethnic origin, unpaid work, and household relationships to include same-sex couples. Finally, many other changes relating to most of the modules in the questionnaire were tested.

Following the NCT, further qualitative testing took place prior to finalization of the questions when they were published in the Canada Gazette in April 2000. The 2001 long form questionnaire contains three questions that were not asked in 1996; they are on religion, birthplace of parents and language at work. Another important change saw two questions on both the short and long questionnaires altered to reflect an expansion of the definition of common-law relationship to include both opposite-sex and same-sex partners. The 2001 Census questions are available from the Statistics Canada Internet site.

2. Data Collection

With the exception of the Internet test, described below, no major changes will take place in the 2001 Census as compared to the 1996 Census. In most areas of the country enumerators will drop off a questionnaire to each dwelling and respondents will mail back completed questionnaires to the enumerator. About two percent of households will be enumerated by interview. This method is used in remote and northern areas of Canada, on most Indian reserves and in parts of the urban cores of some major cities. In addition, some remote northern areas are enumerated in February or March. This advance Census is carried out in areas where communities disperse in the spring and migrate to their hunting and fishing grounds.

A multilingual census help-line is available for four weeks to assist respondents who have difficulty understanding and answering their questionnaires. A Public Communications Program supports the Census by heightening public awareness and by encouraging complete, accurate and punctual responses to the questionnaires. This is conducted via the voluntary support of corporations, associations and governments, as well as through schools, the media and paid advertising.

In preparation for possible changes in the 2006 Census a number of tests are included in the data collection procedures. In one of these, a large scale test will assess alternate edit procedures to identify questionnaires requiring follow-up. As well it will evaluate the efficacy of telephone follow-up with and without secondary follow-up by personal interview. A second and more visible of these will be an Internet test, where people in two selected

places will be offered the option to answer the questionnaire using the Internet.

The Internet offers significant opportunities for data collection and a number of national statistical offices are already collecting some data via the Internet. The government's "Connecting Canadians" initiative fosters the conduct of business with Canadians electronically wherever possible by 2004. Further, Internet connectivity rates in Canada are projected to be in the range of 60-70% by 2006. So, collection via the Internet is a realistic consideration for the 2006 Census. The 2001 Census, given its mandatory nature and high public profile, provides the best opportunity to conduct a test prior to the 2006 Census.

The objectives are: to obtain a measure of public acceptance of electronic alternatives, to evaluate data quality provided via Internet versus that provided by similar respondents on paper, to obtain some measure of informatics technology infrastructure requirements, and to assist in development of a standardized approach to the development of web-based data collection for surveys.

The Internet option will be available for respondents in two test sites with a total of about 200,000 households and to persons with visual disabilities across Canada. For planning it is assumed that the household Internet connectivity level is 35% and that about 20% of such households will use this method of response. An important concern is that of security of respondent data for which procedures are being developed.

3. Geographic Issues

Statistics Canada's digital data holdings for streets and boundaries up to and including 1996 were limited to urban areas where the population exceeded 50,000 people, accounting for 63% of dwellings, 62% of the population and less than 1% of the land area in Canada. Paper maps were used for the rest of the country. Following the 1996 Census, Statistics Canada initiated a project to complete spatial data coverage of the entire country, resulting in the creation of the Spatial Data Infrastructure. It is a large, integrated system of spatial information, such as streets, boundaries and topography, and attribute information such as street names, addresses and a variety of other identifying characteristics. From the perspective of the Census this has facilitated the automated production of all maps required for the Census more cheaply, more easily, more rapidly and with higher quality than in the past.

During the preparation for the 2001 Census, Statistics Canada has developed several tools that have significantly improved the efficiency and efficacy of tasks that were previously done manually. Of particular significance is the Generalized Area Delineation System (GArDS), developed and implemented to automate the creation and delineation of areas for the collection, processing, dissemination and analysis of statistical information. It can optimize area boundaries to meet the specific needs of any application. For the 2001 and following Censuses it will be used to delineate many areas including enumeration, dissemination, weighting and urban areas.

In 2001, all of the Census questionnaires will be geocoded to the block level. This will facilitate the creation of a new area for Canada, the dissemination area. Previously, the enumeration area was used for both collection and dissemination. Dissemination areas will be more consistent in size and more stable over time than enumeration areas and more relevant to users of Census data.

Statistics Canada's Address Register is part of the Spatial Data Infrastructure. It was initially created for the 1991 Census to serve as a coverage improvement tool by providing an alternative listing of addresses that corresponded to enumeration areas. This list was provided to the enumerators after they had performed the standard list/leave (drop off) methodology utilized for the majority of enumeration areas. Discrepancies between the two independent lists are field verified to direct the enumerators to potentially missed dwellings. The Address Register was used in the same fashion for the 1996 Census and will be similarly utilized in 2001. The coverage of the Address Register currently corresponds to urban areas with at least 50,000 population and includes about 63% of dwellings; this coverage will be substantially expanded in the coming years leading up to the 2006 Census.

4. Data Processing

In a quest for improved efficiency, better timeliness and flexibility, Statistics Canada went to great lengths in the 1996 Census to replace traditional approaches to Census processing with a more parallel processing approach. This resulted in less confinement to mainframe computing and more use of PCs and servers for interactive-type applications. The net impact has been time and labor savings as well as paper reduction. The 2001 Census processing strategy will consolidate this approach further.

In 2001 we will again use Canada Customs and Revenue Agency facilities and trained personnel to data capture Census questionnaires. For the first time, we will scan electronically all visitation records and all completed questionnaires as they are received at the 2001 processing site.

The questionnaire and visitation records will then be used in an image retrieval system instead of the paper. This will have a major impact on subsequent operations requiring direct reference to questionnaires and/or visitation records. In addition to a substantial reduction in the use of paper, concurrent access by multiple clerks to the same document will be possible, and inefficiencies and errors related to maintenance and use of a large library of paper documents will be avoided.

The automated coding operation converts textual responses to corresponding numeric codes for a number of variables, excluding industry and occupation. Match rates in excess of 90% have been achieved in the past. By the use of expanded reference libraries from responses to the 1991 and 1996 Censuses and by dynamic updating while doing the 2001 coding it is hoped to reach still higher rates.

A Nearest-Neighbour Imputation (NIM) methodology was developed for the 1996 Census and applied to a number of demographic variables. It offered several advantages compared to the previous software including minimum change imputation of qualitative and numeric variables simultaneously, a reduced tendency to falsely inflate small population groups, applications to larger groups of variables, more plausible imputation actions and superior computational efficiency. The Canadian Census Edit and Imputation System (CANCEIS) implementation of NIM has been improved with added flexibility, portability, and support. For 2001 it will be used for an expanded set of variables. It is planned that CANCEIS be used for all variables in the 2006 Census.

5. Coverage Measurement

The coverage error measurement program, largely unchanged from 1996, measures undercoverage and overcoverage of population and of households. As well, classification errors involving unoccupied private dwellings are measured. Four studies are conducted. The Dwelling Classification Study (formerly known as the Vacancy Check) produces estimates of undercoverage arising from the incorrect classification of dwellings as occupied or unoccupied. It will also facilitate improved imputation for the small percentage of non-responding dwellings. The Reverse Record Check, conducted via Computer Assisted Telephone Interviewing for the first time, measures undercoverage from all sources using frames independent of the current Census. Overcoverage is measured by the Reverse Record Check, the Automated Match Study and the Collective Dwelling Study. The second of these focusses on persons counted more than once within the same region of the country while the last estimates overcoverage resulting from persons enumerated as usual residents in a collective dwelling who were also enumerated at a private dwelling.

6. Dissemination

Census dissemination for the 1996 Census was very successful in its move towards a much stronger orientation to electronic publication. This direction will be pursued further for 2001 with the development of a wider variety of products tailored for different market sectors. A consultation process, similar to that used for content determination, has been undertaken. The release schedule has been planned so that closely related variables will be released at the same time, thereby facilitating a shift in focus towards more analytic content using a thematic approach. The presence of the Census on the Internet will be expanded and coordinated with other Statistics Canada data dissemination initiatives. For example Census Community Profiles, free on the Internet, will be expanded to incorporate data from other sources

and will provide data for non-census geographies. More generally, it is planned to increase the amount and detail of Census data made available free on the Internet.

Some developments in geography will be important for dissemination. First, will be the use of dissemination areas instead of enumeration areas. As well, the concepts of Census Metropolitan Area and Census Agglomeration have been extended to cover the whole country in terms of the degree of urban influence; these will meet the expressed needs of numerous analysts.

7. Post-Censal Surveys

After having two in 1991 and none in 1996, four post-censal surveys are planned for 2001. Two of these, the Aboriginal Peoples Survey and the Health and Activity Limitation Survey, were last conducted following the 1991 Census. The other two are new: the Ethnic Diversity Survey and a survey related to literacy. All four will select all or part of their samples from the 2001 Census with data collection taking place in the year following Census day.

8. The 2006 Census

Strategic planning for the 2006 Census has already started. A number of tests are included in the 2001 Census and planning for the 2004 National Census Test has started. Major changes being assessed include: use of the Address Register to facilitate mailout of Census questionnaires to a large percentage of Canadian dwellings, mailback to centralized locations for edit and follow-up, data capture at centralized facilities via optical scanning and image recognition technology, offer of response via Internet to all households, expanding the use of the Nearest-Neighbour Imputation Methodology to all variables and a number of content alternatives.

9. Conclusion

Although characterized earlier as a Census with incremental changes from 1996, the changes are numerous and collectively represent a substantial challenge. However, they all will contribute to again ensuring a high quality Census and take initial steps towards possible major changes in 2006.

More information is available at the Statistics Canada Internet site, www.statcan.ca or through the office of the 2001 Census Manager: Benoit Laroche, 2001 Census Manager, Statistics Canada, Jean Talon Building, 3rd Floor, Ottawa, Ontario, Canada, K1A 0T6. E-mail: benoit.laroche@statcan.ca.

A COMPARATIVE STUDY ON CHINA'S 2000 POPULATION CENSUS AND THE PAST CENSUSES

Zhang Weimin
Director-General
Department of Population, Social, Science and Technology Statistics
National Bureau of Statistics of P.R. China

Summary

As with many other countries, a new population census, the 5th population census, was conducted in China in the year 2000. This was the first population census during the transition period from a planned economy to a market economy. The result and methodology of the census will exert a great impact on establishing a socialist market economy and on realizing strategic targets for China's modernization.

Since the first population census in 1953 there have been four censuses conducted in China. The Chinese government decided in the 1980s that a population census would be conducted every 10 years in the year with a last digit of 0. The past four population censuses have proved beneficial. The third and the fourth population censuses in particular have been commended by domestic and international experts.[1] All past lessons from working on the census are helpful in ensuring successful implementation of future censuses.

While the past successes can no doubt be taken as the basis of China's 2000 population census, it is impossible to repeat what we did during the cross-

[1] Proceedings of International Seminar on China's 1982 Population Census, compiled by the Office of Population Census of the State Council and the Department of Population Statistics of the National Bureau of Statistics.
Proceeding of International Seminar on China's 1990 Population Census, published by the Statistical Publishing House of China, December 1993.

century population census because China's economic system and population status have changed, and the information-gathering techniques have also developed rapidly. The constant improvement and creation of census methods and technical measures are needed in the light of these changes and development.

This paper provides a brief introduction on the methodological and technical differences between China's 2000 population census and its past censuses. Methods for improvement are also introduced.

1. Changes of Environment for Population Census

The design of population census and the methods adopted during the census are closely linked with environment for population census such as social and economic development of the country and population status, etc. To understand the differences between China's 2000 census and its past censuses, one must first understand the environmental changes of the census.

a. Change of Requirements

The purpose of past population censuses was to find out the number of people and the geographic distribution of the population. Following China's economic development and the improvement in standard of living, the needs of the government and the public have also changed. They want not only to know the number of people and geographic distribution of population but also to understand the quality and structure of the population and the changes of living standards. The year 2000 is a year of special historical significance; the requirement concerning the population's development in a new century was critical for the census.

b. Difficulty in Selecting Enumerators

Before the 1990s a planned economic system had been implemented in China; the socialist market economy had been established. Under a planned economic system, ownership is unique; enterprises basically belong to the government. With this system, it was easy to find enumerators for the census. However, under the market economy, some of the enterprises have been changed into shareholding and private enterprises. It was more difficult to find enumerators from enterprises with various types of ownership.

c. Growth of Floating Population

Before the 1990s there was less floating population. Since the economic reform and especially since the establishment of market economy, great changes have taken place in the aspect of employment mechanism. The floating population substantially increased, especially after the release of the large amount of rural surplus labor. This activity made it more difficult to have an accurate population count.

d. Reduced Household Registration

The perfect administration of household registration in China was the basis of past censuses. But since the economic reform and the establishment of a market economy, people have less interest in registering their place of residence. The population who have not registered their household has increased, and there are more and more people with inconsistent residence and household registration. The administration of household registration has, unfortunately not adapted to the change of requirement. Therefore, it is no longer possible to rely wholly on household registration to conduct the population census.

e. Reduced Citizen Cooperation

Following the growth of the market economy in recent years, the sense of individual privacy has been strengthened. In addition, people have become disgusted with the frequent commercial surveys, and they are not willing to answer questions. This behavior influenced citizen cooperation during population census.

2. Organization of the Censuses

Since the population census is a massive statistical activity that relates to every household, it is essential to set up an authoritative agency to conduct the census. Otherwise, it will be difficult to mobilize the public to participate in the census. In the past, a leading office for the census was set up by the government at different levels (central, provincial, prefecture, county, and township levels). The heads of the leading offices were usually the vice premier, vice governors of provinces, etc. The

members of the leading offices were usually the heads from various departments, such as statistical, planning, financial, public security, propaganda, civil affairs, family planning agencies with the responsibility of organizing, fundraising, and enumerator-recruiting. Under the leading office of the census, an administrative office was set up consisting of the members from statistical, public security, and propaganda agencies and responsible for practical activities related to technical activity and implementation of the census. This type of organization proved effective and necessary under the current level of social and economic development in China. Such organization was also set up in the 2000 population census.

During the 2000 population census, the enumerators were selected mainly among governmental staff, teachers of primary and middle schools, and cadres of village committee. Compared with past censuses, the enumerators from enterprises were reduced.

3. Field Enumeration

Field enumeration was used during the past four population censuses of China (i.e., the enumerators were responsible for filling in the questionnaires). A large number of enumerating stations were set up during the first and second population censuses. Respondents were requested to report the status of household members in the appointed enumerating stations, and then the questionnaires were filled in by enumerators. Although enumerating stations save time, they compromise individual privacy. Therefore, household enumeration was adopted during the third and fourth population censuses. Enumerators visited each household and filled in census questionnaires. During the 2000 population census, this method was also used.

In many developed countries, household enumeration is used (i.e., the questionnaire is filled in by the member of household themselves). Since China is a developing country with a large proportion of agriculture population, the ratio of illiterate population reaches 12 percent, it was difficult for members of households to fill in questionnaires by themselves. For this reason, the household enumeration was not suitable.

4. Respondents

During the population census, a clear definition was made on the place where the respondents were enumerated in order to avoid underreporting and overreporting.

During the past four population censuses, the principle of permanent residency was adopted for enumeration (i.e., the respondents were enumerated at their permanent residence). Certainly in many countries, the concept of current residency is used during the census (i.e., all residents are enumerated at their place of residence on the day of the census). But it is now difficult to do so because if we choose such a method, the field enumeration time would be reduced (e.g., within one day) in order to avoid the influence of the floating population on the enumeration. By using household enumeration we would have to increase the number of enumerators from 6 million to 12 million or 24 million. Obviously, we cannot bear such financial and organizational burden. Therefore, the concept of permanent residency was adopted during the 2000 population census.

Compared with past population censuses, although the respondents are permanent residents, the time and space for permanent residency were adjusted during the 2000 census.

a. The time for permanent residency was reduced from 1 year to 6 months. Following the active development of the market economy and the changes in employment, the residential status items will change more frequently. The reduction in the time for permanent residency has just adapted to such changes. In addition, there is a traditional custom in China where people will go back home for the Spring Festival, especially those farmers who are engaged in business in cities, and stay 1 or 2 months. If the time for permanent residency was still 1 year, many farmers who have been in cities for many years would inevitably be enumerated in rural area, and their occupation would still be reported as farmers. The changing process of tertiary development and increase of floating population since the reform and opening to the outside world would be concealed.

b. The space for permanent residency was reduced from county (city) to township (street). According to the administrative structure of China, the administrative scope of county (city) covers both the township of rural areas and the town of urban areas. In recent years, significant rural surplus labor has floated not only to the developed large cities but also to small towns nearby. If the space for permanent residency is defined as county (city), those people would still be enumerated in their rural residence and the status of the floating population would not be reflected.

5. Reference Time

During the past four population censuses of China, the census was set at zero o'clock of 1 July (i.e., 1 July is the census day). The reference time of the 2000 population census was changed to 1 November for the following reasons:

a. During November, there is no large amount of floating population.

b. China has a vast area of land, and its climate is quite different in southern and northern areas. The period between June and July is not only the hottest but also the rainiest season; floods often occur. As most levels of governments and residents are usually engaged in flood control and salvage, it is difficult to conduct the census. It is also not convenient to conduct the census during the winter because of the cold weather in the northern area. With regard to rainfall and temperature, April, May, October, and first half of November are the most suitable months for the census.[2]

c. Considering the utility of the census data, the time for permanent residency was set for 6 months. If the reference time of the census is set in April or May, since many farmers return back to the cities within 6 months after the Spring Festival (usually in February), they may still be enumerated at home. The status of the floating population cannot be reflected by the census data. If the reference time of the census is set at the first half of November and the 6-month period refers to the period from May to October, this problem can be solved.

6. Census Items

The number of items on the census was generally smaller in the past censuses, and the items usually focused on the changes within the population. They did not reflect the economic activities of the population, standard of living, or quality of life and did not relate with housing indicators.

The items of the 2000 population census were not only increased but also combined with items on housing for the first time. The items of the 2000 census include the following:

a. Individual (22 items)

- Basic items: name, relation to head of household, gender, date of birth, nationality, etc.
- Migration items: place of household registration, place of birth, permanent residency 5 years ago, period of current residence, reason for migration, etc.
- Educational items: ability to read and write, educational level, schools, etc.
- Economic activity: employment status last week, time of work, industry, occupation, unemployment, job-finding, etc.
- Marriage and births: marriage status, length of first marriage, number of children ever born, number of living children, births in last 12 months, etc.

b. Household (23 items)

- Basic items: type of household, number of household members, births, deaths, number of household members who were absent for more than 6 months, etc.
- Housing items: floor space, number of rooms, date of construction, materials used, kitchen, toilet, water, electricity, fuel, bathing facilities, original ownership of the house, monthly rent (expenses of purchasing the house), etc.

c. Deaths (8 items)

The items about the deaths cover some basic information of the people who died during the period from 1 November 1999 to 31 October 2000

[2] A Preliminary Study on the Suitable Season for Population Census, written by Lin Zhiguang. Paper from the Symposium on China's Population Census, November 1993, Beijing.

such as name, gender, date of birth, nationalities, etc.

Compared with the 1990 population census, the new items added mainly include birth place, time of residency, time of work last week, source of income, and all items about the house. During the past population censuses, items about the house have never been included. It was one of the characteristics of the fifth population census added to the items about the house.

7. Questionnaires

The content and number of items of the 2000 census increased to a great extent compared with past ones. If everyone was enumerated with all the items, the workload of enumerators would become much heavier, and the financial and material input would also increase. To solve this problem, a short questionnaire and a long questionnaire were used during the 2000 census. With this method, we could get more information and reduce the input. The short questionnaire consists of basic items, including type of household, number of household members, number of rooms and floor space, name, relation with head of household, gender, date of birth, status of household registration, nationality, educational level, etc. Ten percent of households were selected to fill in the long questionnaire for all items.

Many countries use short and long questionnaires during their census and have accumulated substantial experience in applying such techniques. Since this was the first time to use this technique in China, detailed and concrete rules and instructions were drawn up to address many issues, such as the selection of the sample and calculation of the sampling error and field operation to secure success.

8. Data Entry, Data Processing, and Data Dissemination

a. Data Entry

Keyboard typing was used for data entry during China's 1982 and 1990 population census.[3] While there were some advantages by using this method, there were also some disadvantages. First, more operators were needed and the cost of staffing increased. Under the planned economy, the cost of staffing was very low; the data entry could be completed with little or no cost, which cannot be done under the market economy. Second, data entry by keyboard typing takes a long period of time. The results of modern population census should be provided to the users as early as possible in order to put the role of the census into full play. Third, during the data entry, typists have to use their eyes to identify figures and use their hands to type figures. It is much easier to make errors; the quality of census data is affected.

As the cost of labor has increased in recent years and the government at different levels and the public hope to get the census results within a short period of time, Optical Character Recognition (OCR) was used for data entry during China's 2000 census. Since it is the first time to use OCR in China, great attention was paid to the quality and accuracy of printing the questionnaires, quality control measures, and the standardization of filling in questionnaires by enumerators to guarantee the success.

b. Data Processing

The data entry and data processing of the past censuses were conducted in a decentralized way. During the 1990 population census, data entry, cross-check, and the tabulation of the county and prefecture census results were conducted by 380 prefecture census offices. The processed data were kept on floppy disks and sent to the provincial census offices where checks and provincial tabulations were made. The processed data of provinces were also kept on floppy disks and sent to the national census office where checks and national data processing were conducted. There were obvious advantages to using the decentralized method for data processing. First, the time for data processing was reduced; the local governments and data users could get the census results much earlier. Second, it was convenient to check and correct the errors occurred during data processing. Such a method of data processing proved to be successful.

[3] The Design and Implementation of Data-processing System of China's 1990 Population Census, P1-3, issued by the Computer Center, National Bureau of Statistics.

c. Data Dissemination

The extensive use of census results is one of the signs indicating the success of the census. In the past, census data were disseminated in paper-printed publications. Along with the development of the computer and computer networks and the growth of the market economy, it was more difficult to meet the need of development to fully follow the traditional ways of data dissemination. Improvements in this task were made during the 2000 census of China.

- The application of geographic information system (GIS)
  GIS has never been used in the past population censuses of China. To raise the technical level of the population census and extend the use of the census results, GIS was applied in large and medium cities to achieve some experiences of construction and application and lay a strong basis for setting up GIS all over China in 2001. GIS will not be applied in rural area for a period of time.

- The application of computer and network transmission techniques
  In the past some limitations existed in disseminating census data through paper-printed publications in the past. During the 2000 census, the paper-printed publications were improved. The important aggregated and the public-needed census data was printed and published by combining figures, pictures, and articles, and is easy to understand. Such publications can not only be used for widespread dissemination, but also to popularize the census information and strengthen people's awareness of population. The detailed and specialized census data was published using a magnetic media such as floppy and optical disks, or provided to the users through computer networks.

9. Conclusion

Many successful and proven experiences and methods from the past four censuses were used in China's 2000 census. Some of the methods of the past censuses, such as the establishment of an authorized leading office, social mobilization, the method of field enumeration, and the ways of data processing were kept.

As for the number of census items, related methods, and techniques, further improvements were made in the light of the market economy and international practices, such as the newly added items about housing and people's life, newly revised items of migration, economic activities, and education, the use of long and short questionnaires, the application of OCR and GIS techniques, and data dissemination. The 2000 census of China is a census based on China's traditional method combined with advanced international practices.

Bibliography

Census Office of the State Council and the Department of Population Statistics. (March 1984). *Proceedings of the International Seminar on China's 1982 Population Census*. Beijing. National Bureau of Statistics of China.

Census Office of the State Council and the Department of Population Statistics. (October 1992). *Proceedings of the International Seminar on China's 1990 Population Census*. Beijing. National Bureau of Statistics of China.

China's Association of Population. (1997). *Population Forum of China-23rd International Population Conference*. Beijing.

Department of Population Statistics, National Bureau of Statistics of China. (October 1995). *Scientific Research on the Method of China's Population Census*. Statistical Publishing House of China.

Li Chenri. (November 1987). *Analysis on the Results of China's Population Census*. Financial and Economic Publishing House of China.

Qiao Xiaochun. (October 1995). *Study on China's Population Census*. Population Publishing House of China.

Suggestions and Principle of Population and Housing Census. UN Secretariat, ST/ESA/STAT/SER.M/67/Rev.1.

Yu Y.C. (March 1997). *Method of Collecting Population Data*. Statistical Publishing House of China.

Zhang Weimin. (January 1995). *A Test under the Market Economy- Introduction on 1995 1% Sampling Survey of China*. Market and Population Analysis.

Zhang Weimin. (1995). *The Application of Sampling Method in the Data-processing of China's 1982 Population Census*.ESCAP Seminar on 1990 Population Census STAT/WPHCP/CPR.1,ESCAP.

Zhang Weimin. (1996). *Enumeration Method and Quality Control of Population Census*. UN Expert Meeting on Population and Housing Census, New York.

Zhang Weimin. (1998). *On the Issues of China's 2000 Population*. Population and Economy, Vol. 4, Capital University of Trade and Economy.

Zhang Weimin, Cui Hongyian. (October 1992). *Evaluation on the Data Quality of China's 1990 Population Census*. International Seminar on China's 1990 Population Census, Beijing.

Zhang Weimin, Yu Hongweng, Cui Hongyan. (October 1997). *Current State of Changes of China's Population*. 23rd International Population Conference, Beijing.

Zheng Jiaheng, Zhang Weimin. (May 1992). *The Application of the Data of China's 1990 Population Census*. 14th ESCAP Seminar on Population Census, Tokyo.

FINLAND: NEW STATISTICS MADE POSSIBLE BY THE USE OF REGISTERS[1]

Pekka Myrskylä
Statistics Finland
Population Statistics
Email: **pekka.myrskyla@stat.fi**
Tel. +358 0 1734 3555

The use of administrative records in Finnish population censuses began in 1970 and has been increasing ever since. The Population and Housing Census of 1990 was the first totally register-based, without using a single questionnaire. Since 1987 Statistics Finland has produced an annual database for the entire population. Information is produced for the nation but can also be produced for smaller areas such as provinces, municipalities, and parts of municipalities.

Census data are used to compile longitudinal data files in which each resident of the country is linked with his or her data in different censuses. And the other data file combines the annual data since 1987. This makes it possible to follow the life course of each person. Changes in the labour force has been compiled by monitoring the flows of population and labour force from one status to another in different years.

1. Introduction

Only a few countries in the world such as Denmark and Finland successfully use population registers supplemented with other administrative records to take their censuses. Sweden, Norway, and Iceland are following this lead in the near future. Also, in Central Europe some countries are using more and more administrative records in the production of data.

There are many attractive features to a register-based census system. It is relatively cheap and efficient to use existing administrative records to develop information. There are, however, tradeoffs in terms of the types of information that can be collected and the quality of the information. In our experience, timeliness of data production remains about the same as with a traditional, questionnaire-based census.

We do not worry about using time questionnaire design, how and in what order questions are asked, or about producing questionnaires in multiple languages or writing questions so the barely literate can respond. In short, a register-based census allows the central statistical office to concentrate on what they do best—statistics.

The register-based census system is actually **more secure** than the systems from earlier censuses because (1) we do not need to rely on **outside mailing** and **recording service**s (2) there is much **less manual treatment** of personal data, and (3) **full personal data is only "seen" by the computer** because data are processed in batch runs and the files are sequential files.

Methodology needs to be developed to estimate data quality, including how to estimate and present statistical error in administrative records and databases. This has been the subject of world-wide discussions. Work so far has been limited to consistency over time within a database and comparisons of different databases. There is nothing analogous to the concept of "standard errors" as in surveys.

[1] This paper was first presented in Joint (ECE/Eurostat) Work Session on Registers Administrative Records for Social and Demographic Statistics (Geneva, 1-3 March 1999).

2. Advantages of the Register-based Statistical System

- This system allows for **Lower costs, no additional burden** to respondents, the additional **use of existing register data,** and many new statistical possibilities.
- **The system provides for the annual availability of total data** for the country and **small geographic areas** and small population subgroups.
- The **register system captures short spells of activities** and **multiple activities** or side activities such as employment among students, double or triple simultaneous working relationships, etc.
- There is also improvement in data quality for items that would otherwise involve **long-term memory** or detailed records.
- **The coverage of register-based data system is usually higher**. We have significant supporting information, for example, in coding occupations, all the previous codings of individuals, educational qualifications, data describing employers (the size and industry employers).
- The system provides the ability to **link various types of data sets** such as demographic and health characteristics or real demographic and business information.

3. Disadvantages to a Register-based Census

Some subjects are **not available** on administrative records.

a. In a traditional questionnaire-based census, there is more **flexibility to ask questions on emerging issues**.
b. Because the data are not available in registers, the 1990 census of Finland does not include information on **hours worked during a week**. This information should be essential to define part-time and full-time workers.
c. There is no information on the ways to **commute to work** because no register covers such topics.
d. There is no register data for people **engaged in housekeeping,** and so they must be categorised as other economical non-active groups, as the residual of all other groups.
e. There is no information on the **ethnicity** of the population in Finland.
f. There is no information on the **attitudes of the general public**.

Perhaps the greatest danger of all to the use of administrative data for statistical purposes is posed by the attitudes of the general public. The use of such material and the possibilities for combining the contents of extensive registers of this kind can arouse a fear of infringement of personal privacy in some sectors of the population, as the authorities are in possession of the "whole picture" of the life history of individual citizens (cf. longitudinal data file from population censuses).

Frequently, however, these sentiments are linked chiefly with a fear of administrative action of the "big brother" type against citizens, which would be utterly impossible using the kind of data that a statistical office possesses. Such speculation has scarcely ever been put forward in Finland, not even in the context of national population censuses. When Statistics Finland has distributed information on its working methods to the press at the time of performing a register-based population census, the publicity has almost always been favourable. The cost savings (US$ 20-25 million) and the convenience from not having to ask the public to fill in forms with information that has already been provided to some government office has been regarded as such obvious advantage that no serious discussion has arisen regarding questions of confidentiality.

Despite the disadvantages, the advantages of administrative records, especially relatively lower costs and the ability to obtain frequent, small area data, are encouragements for statistical agencies to examine their use more closely. This paper examines Finland's experience with census-taking by registers. For the United States, we examine the limited experience with administrative records, the barriers to more extensive use, and new opportunities the United States statistical system is considering.

4. Cost of Doing a Register or Records-based Census

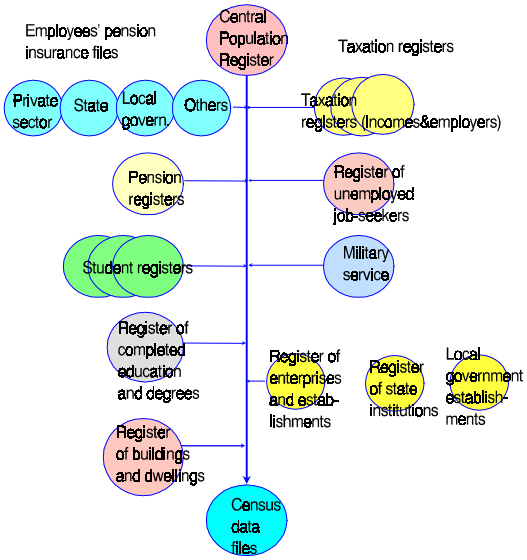
The drop in the cost of doing a records-based census compared with a traditional questionnaire-based census is dramatic. The 1980 census cost about US$ 23 million. The 1990 census cost about US$ 3.5 million, including the expenses of a very

comprehensive quality survey. The comparison is in constant dollars. Thus, the 1990 census cost 15 percent of the census 10 years earlier. Additionally, the number of staff needed also fell considerably, from 650 person years in the 1980 census to about 30 person years in 1990. The costs of the 1995 census was about US$ 1 million, without any quality survey.

5. Register-based System

Data are linked among persons, incomes, employment, unemployment, pension periods, buildings, dwellings, enterprises, and workplaces (establishments). Geography is determined from building co-ordinates so sub-areal data can be produced. In the population registers, each child's record carries the identification numbers of his or her mother and father; likewise, the identification number of each child is included in both his or her mother's and father's records. With this cross-recording of identifications, family members can be matched even if they are not living in the same household. Adult children can be linked to elderly parents living in different households, and researchers can determine if elderly living alone have family nearby. It is possible to combine background data (family data) with students or unemployed.

Figure 1. Use of registers and administrative records in the register based statistics



6. New Research Opportunities

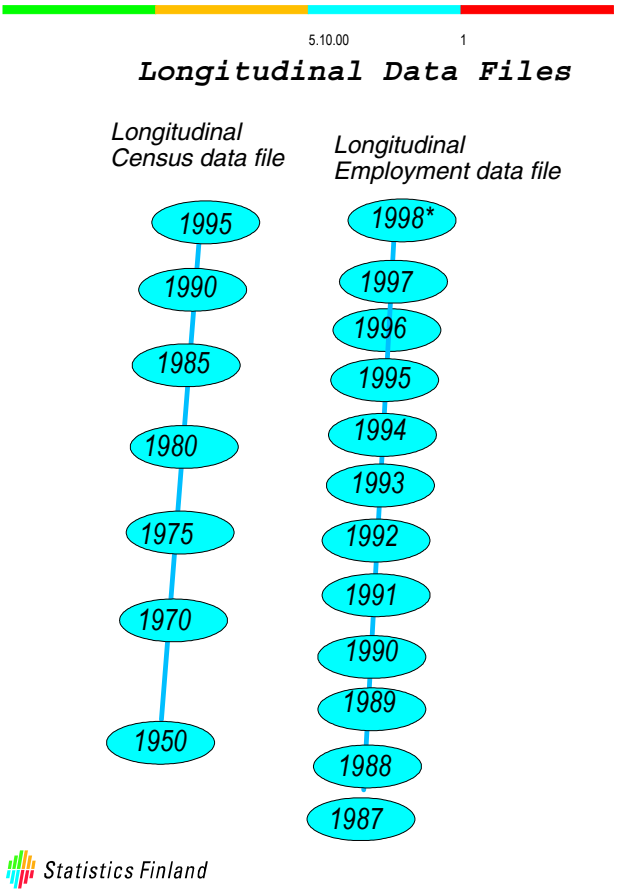
As discussed above, we use our population register and administrative records as Finland has produced an annual employment database for the **entire population**. Information is produced on basic data in place of traditional data collection systems. We have also developed new, cost-effective opportunities for data development and evaluation.

The annual statistics cover all the population and family statistics, employment statistics, building and dwelling statistics, housing statistics, and educational statistics. Since 1987 statistics produced for the nation but can also be produced for all **small areas,** such as provinces, municipalities, and parts of municipalities.

The database provides **longitudinal data**. It is possible to study, for example, changes in occupation and the place of residence and work and related characteristics, such as unemployment, receipt of pensions, and disability. With the database, Statistics Finland can determine how various age groups fare in the labour market and how their work compares with their educational attainment.

Census data are used to compile longitudinal data files in which each resident of the country is linked with his or her data in different censuses, (1970, 1975, 1980, 1985, 1990, and 1995). All these censuses have used register-based data. From the 1950 census we have picked a 10 percent sample, which is linked with later censuses. And the other data file combines the annual data from 1987-1997 as "**short longitudinal data file"**, but the data content of it is much more comprehensive. This makes it possible to follow the life course of each person and the changes that have occurred during a period of 25 years and even as long as forty-five years.

Our registers are used extensively to **select samples** for surveys. All the register units as well as addresses are updated continuously. Registers can also be used to **select samples for specific populations** such as age, income groups, or specific geographical areas. Another use of registers is to **reduce the number of questions respondents are asked** in sample surveys.



For example, the Finnish Labour Force Survey uses addresses, demographic data, educational data as well as family data (family and household combination) already collected in the registers to reduce the number of questions asked. Another use is to determine the characteristics of survey **nonrespondents** (e.g., age, gender, employment status, and place of residence).

FRANCE: SAMPLING, ESTIMATION AND PRECISION IN THE REDESIGNED CENSUS OF POPULATION

Jean Dumais, INSEE, Rhône-Alpes

1. Introduction

During the last decade, the governments of many Western countries cut their budgets. In some of those countries, the national statistics agencies are finding it increasingly difficult to justify and secure the resources needed for a full enumeration of the population. Moreover, in a number of European nations, the average citizen is having more and more trouble understanding why the decennial census is needed when government agencies hold a wealth of administrative data. In light of these environmental changes, France has decided to take a completely fresh look at how it conducts its population census.

One of the objectives of this redesign is to be able to maintain the census tradition of publishing results for every level of detail (even the subcommunal level), of course subject to confidentiality restrictions. With administrative data and modelling, it will be possible to achieve this objective every year.

2. Sampling Strategy

The commune is the linchpin of the redesign effort. For "small communes" (population under 10,000), a 20% sample (on average) will be surveyed each year, and every dwelling will be visited. All "large communes" will be visited once a year, but only a fraction (about 1/12) of their dwellings will be surveyed. Since the population is just about evenly split between small and large communes, the average annual sample will be roughly 1/7 of the dwellings.

a. Small Communes

We will look at the universe of small communes first. In each of metropolitan France's 22 regions, the small communes will be assembled into five rotation groups. These groups will be formed, using information from the March 1999 census, by balanced sampling (Deville & Tillé, 1999) of the counts of the communes' age-sex distributions. The balance variables are the age-sex group counts from the 1999 census; the age ranges are 0-19, 20-39, 40-59, 60-74, and 75 and over. This approach should minimize the year-to-year variations due to sampling alone.

The balance was tested for stability in the small communes of the Rhône-Alpes region (2,811 small communes) using 1982 and 1990 census data. The tests showed little or no loss of precision over the eight-year interval.

b. Large Communes

The census of large communes will be based on areas called "îlots regroupés selon des indicateurs

statistiques" (IRISes) (blocks based on statistical indicators), each with a population of about 2,000 to 2,500. Each year, every IRIS in every large commune will be surveyed for the census. Within each IRIS, five rotation groups of buildings will be formed using the same sampling model as in small communes. The sample frame will be kept up to date by using the "répertoire d'immeubles localisés" (RIL) (inventory of located buildings). This inventory is a list of addressed buildings (residential, institutional or commercial) geographically located so as to generate a digitized map. Initially, the RIL will be populated with data from the 1999 census, which will provide a statistical portrait of each residential building. Subsequently, it will be continuously updated using building permits, demolition permits, utility records (water, gas, hydro, etc.), information supplied by local governments, and field observations. Thus, the RIL can be used to form a building sample frame for large communes.

Three additional strata will be formed in each IRIS: one for industrial buildings (factories, warehouses, etc.), another for collective dwellings (institutions, group homes, communal groups, boarding schools, etc.) and a third for new buildings. One fifth of the industrial buildings will be visited each year to verify that they contain no dwellings (custodian's quarters or space converted for habitation); any dwellings found in such buildings will be considered self-representing because of their special nature (the buildings are not supposed to contain any living quarters). All collective dwellings will be covered each year; 20% of them will be visited, and the population counts of the remaining 80% will be updated, possibly by telephone. Finally, all new residential buildings will be enumerated so that they can be placed in a rotation group and their statistical characteristics can be recorded.

As noted above, each building rotation group will be visited once in each five-year period. Each year, a list of the dwellings in the residential buildings in the current rotation group will be prepared, and a 40% sample of those dwellings will be surveyed.

The annual sample generated by this design will consist of some 8 million individual forms, 6 million from households in small communes and 2 million from households in large communes.

c. Comparison with Other Designs

Under the sample design for large communes, five rotation groups of buildings are formed by balanced sampling by demographic variables. The effectiveness of this design was compared with that of five competing designs with the same sampling rate: a simple random sample of 8% of the dwellings in an IRIS, a two-phase non-balanced sample, a two-stage sample for which the sample size had to be adjusted upward, and two two-stage samples of clusters of contiguous dwellings. The IRISes of the commune of Sedan (population about 20,500 in the 1999 census) were used for the comparison. The design effect relative to the proposed design ranged from about 40 to over 1,000, depending on the design for the commune of Sedan and for each of its IRISes (Dumais, 2000). At this time, there are no plans to publish the raw data from the annual surveys of large communes, except for high-level aggregations.

3. Collection of Data and Questionnaires

The preferred option at present is to collect data in the January-February time frame, for publication referenced to January 1 of the same year. The collection campaign should not take more than four weeks in small communes and six weeks in large communes.

As in previous censuses, the drop-off/pick-up method will be used. The possibility of allowing respondents to mail their questionnaires back to a central point from the outset is currently under consideration. To ensure the success of this option, which relies on action on the part of respondents, the questionnaire will have to be readable and easy to follow. A number of adjustments will have to be made in the questionnaires to measure migration (UN, 1990) and to track double counting.

4. Direct Estimation and Synthetic Estimation

The annual estimates for large communes will probably be produced by dilatation. In the case of small communes, estimates can be produced only for those communes in which a census was conducted during the year.

Current dissemination plans are to release the national results of the survey conducted early in

year Y on December 31 of year Y. The results for large areas, which have yet to be defined but are expected to have a population of more than 1 million, will probably be published on the same day. The data for all of the small communes visited during the collection campaign will also be released.

Detailed results for year Y-2 will be issued on December 31 of year Y; they will be a combination of survey or census data and synthetic data. The synthetic data will be based on the relationship between observed data and administrative data for a given point at a given time. Current plans, subject to the approval of the Commission Nationale de l'Informatique et des Libertés, are to exploit administrative records at a sufficiently detailed geographic level (building, block) to obtain information about individuals (age and sex according to health insurance files) or their dwellings (property tax files).

When this system becomes permanent, we will have, for small communes surveyed in year Y-5 and year Y (see table below), measurements of variables for individuals (age, sex, industry, occupation, etc.) and dwellings (household size, number of rooms, tenure, etc.) at two different times. Complementary information will be derived from administrative files at a sufficiently detailed level. Then we will be able to measure the bias between the observed data and the administrative data for similar objects (buildings, blocks, etc.). And from this measurement of bias for well-defined aggregates we will be able to derive an adjustment factor that can be applied to administrative data so that the adjusted administrative totals closely match the census counts or estimates.

Rotation group	Reference year of release						Current year
	Y-6	Y-5	Y-4	Y-3	Y-2	Y-1	Y
GROUP I			C		S		S
GROUP II				C	S		S
GROUP III					C	S	S
GROUP IV	C				S	C	S
GROUP V		C			S		C

C = Census data

S = Synthetic estimate

In addition, the synthetic estimates for years Y-4 and Y-1 might benefit from the data collected in year Y; in other words, the adjustment factors based on the most recent census could be computed and projected back over the intercensal period. Then the forward and backward projected series could be combined into a single composite series.

Synthetic projection estimation was tested on the small communes of Rhône-Alpes, for which the rotation groups, 1990 property tax data and 1990 population census data are available. Regression models (Särndal, 1990) were used to estimate the additional error due to massif imputation of the household or dwelling variables; the regressor is the number of principal residences according to 1990 property tax records. Preliminary results show great differences in the relative error, varying from 0.32% for the total number of principal residences to more than 150% for the number of one-room principal residences.

5. Work in Progress

Work is under way on questionnaire design and testing, operational collection methods, effectiveness of imputation plans, data adjustment, and imputation of partial non-response. At this time, plans call for the redesigned census to collect its first data in early 2003.

References

Deville, J.-C., Tillé, Y. (1999), *Balanced Sampling by Means of the Cube Method*, CREST-ENSAI, document interne.

Dumais, J. (2000), " Quelques aspects méthodologiques du recensement rénové de la population en France ", *Deuxième colloque francophone sur les sondages*, Bruxelles.

Särndal, C. E. (1990), " Méthodes pour estimer la précision des estimations d'enquête lorsqu'il y a eu imputation ", *Recueil du Symposium 90 de Statistique Canada : Mesure et amélioration de la qualité des données*, Ottawa.

UN (1990), Principes et recommandations complémentaires concernant les recensements de la population et de l'habitat, Etudes statistiques, ST/ESA/STA/sérieM/67, New York.

PRELIMINARY DESCRIPTION OF THE 2001 HUNGARIAN CENSUS

In Hungary, as in other European and developed countries, significant importance is attached to successfully implementing full-scale decennial population and housing censuses. (The first official one goes back to 1870.) Long and thorough preparatory work comprising legislation, field activities, and data processing have been needed to carry out this huge operation that requires enormous personal, technical, and material resources. On the eve of the new millenary, the importance of preparation is stressed by well-defined, up-to-date data needs, traditions, aspects of temporal and international comparison, and the necessity of finding solutions to financial, human resources, and legal limitations. Suitable preparation is also important due to basic socioeconomic and political changes in recent years, expected data needs of the post-censal period, and new challenges of technical development.

According to our Census Act (December 1999), the forthcoming population and housing census in Hungary will be carried out in February 2001. Taking into account our experience and legal limitation in the use of administrative data sources, we will carry out a traditional data collection based on direct (door-to-door) interviews with questionnaires and enumerators. The number of these latter amount to about 40,000 (i.e., the enumeration districts contain 80-100 housing units with 200-300 persons on the average). Important modifications were implemented (1) in the geographic preparation, (2) on the census forms both from structure and content points of view, (3) in PR activities preceding and accompanying the census itself, and (4) in data processing and dissemination technology. Some of these are as follows:

The preparation of up-to-date address lists — the basis of field preparation — has been computerized and checked in a pilot census in 1999.

Besides the traditional and compulsory sets of demographic, educational, occupational, family, household, and housing topics, the question of legal nationality (citizenship) and that of religion were reinstated on the questionnaire, as well as a block on disabilities. According to our legislation on gathering statistics, these latter topics and those on ethnic nationality and mother tongue are "sensitive characteristics," that is, their inclusion in the questionnaire required careful legal consideration (e.g., the respondent's name is not asked for in the questionnaire and the option not to answer these questions).

The structure of the census questionnaires has been adapted to the new data entry system using an optical reader. In the course of preparing the forms for this process, we were inspired by the way other countries had handled this task. For this process to take place for our census, we encountered several new challenges, such as hiring a firm to scan several million multisided questionnaires.

Likewise for the preparatory phase of the preceding census round, our office took an active part in the compilation of the new year 2000 set of United Nations and European Union recommendations. New occupational and industrial classifications were adopted in the 1990s in conformity with relevant international standards (ISCO-88, NACE).

THE PHILIPPINES CENSUS 2000

The latest Philippine census placed the population of the country as of May 1, 2000, at 75.33 million. This figure is higher by 6.71 million from the 1995 population.

The 2000 Census of Population and Housing, more popularly known as Census 2000, is the latest census of population and housing units all over the Philippines. It is the 11th census on population and the 5th census on housing in the country since 1903. Census 2000 began last May 1, 2000, and lasted for about 1 month in almost all provinces nationwide. The census was completed after 2 months in Metro Manila, the country's premier metropolitan area, and the provinces of Lanao del Sur and Maguindanao, where there is a clash between the military and Muslim rebel groups. The enumeration utilized around 45,000 public school teachers who did the house-to-house visits and interviews in the 42,000 *barangays* (smallest political subdivision in the Philippines) nationwide.

Thinking of the big task ahead, the NSO coordinated and established partnerships with government and nongovernment agencies and individuals. This was seen as an approach that will involve the general public, policymakers, academe, politicians, various sectors of society, and stakeholders, directly or indirectly, in the census. How people and organizations/agencies will recognize the importance and the benefits that can be derived from the activity and take on active roles and consequently establish ownership in the census were the challenges that the NSO worked on. The NSO believed that in establishing the people's ownership in the census, dissemination and usage of results would become a matter of favorable consequence.

A series of consultations and meetings with the data users or stakeholders were vital to census planning. The NSO requested the House of Representatives Committee on Population and Family Relations to sponsor a public hearing with the following objectives: (a) to recommend the concept on household membership to be adopted; (b) to discuss the definitions and categories to be utilized for urban and rural classification; ethnicity and disability; and (c) to come up with a list of items to be included in the census questionnaire. The committee invited several government agencies and the academe/research organizations to two separate committee hearings. There were about 18 government agencies and universities/organizations that participated during the hearings.

A series of meetings was also organized to clarify concepts and categories to be utilized for the census taking to best respond to the needs of special sectors. While the public hearing also included such agencies, a second meeting with these specific agencies like the National Commission on the Indigenous Population, National Commission on the Welfare of Disabled Persons, and Philippine Council for Evangelical Churches helped in the finalization of categories for ethnicity, the definition and coverage of disability, and the further classification of the Protestant category for religion, respectively.

Considering that the census is a politically-sensitive issue, as municipalities and *barangays* contest statistics of population that serve as basis for their internal revenue allotment (budget allocation provided to these municipalities), coordination and partnership efforts with local government units was an advantage. National, provincial, municipal, and city census coordinating councils were established as provided under Section 3 of the Philippine Census Law (*Batas Pambansa 72*). Section 4 of the same law provides for the composition of these boards. The provincial governor and city/municipal mayors act as chairpersons of the provincial and city/municipal census boards, respectively, while the division/city superintendents of schools as vice chairpersons. NSO provincial and district officers act as the executive officers in their respective census boards. Members include officers from various offices in the local government units (LGUs) and agencies. The direct involvement and roles of local officials in the census boards increased their support in the undertaking. It will also provide the census final results an implicit imprimatur (endorsement) from the LGUs because after all they were on top of the census preparations.

The NSO likewise coordinated and mobilized various national and local government agencies, as well as universities, media groups, private organizations, and individuals to get them involved in Census 2000. A major role was played by the Department of Education, Culture and Sports (DECS) because public school teachers were utilized as census enumerators and area supervisors nationwide. While this is mandated by *Batas Pambansa* 72, DECS teachers have, since 1980, given their full support to censuses by painstakingly enumerating their assignment areas. Other forms of assistance received by NSO from other agencies, organizations, and individuals included, but not limited to, the use of their facilities as training venues; use of delivery vehicles to transport the survey materials to/from municipalities and *barangays*; provision of gasoline; use of communication facilities; publicity and promotions (sponsoring newspaper and television advertisements, discussion of Census 2000 events and ceremonies, provision of T-shirts, caps to teacher enumerators and supervisors, etc.), provision of additional manpower as trainers and supervisors.

One of the more important aspects of Census 2000 in the Philippines was the conduct of an extensive publicity and information campaign. For this, the NSO established a partnership with the University of the Philippines College of Mass Communications Foundation, Incorporated (UPCMCFI) in its

information, education, and communication (IEC) strategy for Census 2000. NSO thought of tapping an expert in the field of IEC because an extensive and effective information and communication drive is necessary to publicize the huge undertaking and reach, if not all, the greater majority of the population, considering its diversity and geography, among others. The UPCMCFI came up with a publicity plan for Census 2000 based on input from the NSO. At the national level, the NSO identified and selected target groups to which television/radio/print ads and information campaign will focus. The target groups included the youth/general public, the ethnic population in Mindanao, Filipinos of Chinese descent, and the upper income class or the elite group. Endorsers for each target group came out in television and print ads while their voices were likewise heard over the airwaves through the radio plugs. Senator Aquilino Pimentel, Ms. Teresita Ang See (the spokesperson of the Citizen's Action Against Crime, a civic nongovernment organization fighting against crime like kidnapping) and Mr. Guillermo Luz (executive director of the Makati Business Club) were requested to endorse the advertisements intended for the Mindanao region, Filipinos of Chinese descent, and the upper income group, respectively. A mascot, named as C2K, was conceptualized and created to endorse the ad intended for the youth and the general public. Furthermore, as practiced even in the past censuses, the President of the Philippines was also tapped to endorse the census through a TV advertisement. His Excellency Joseph Ejercito Estrada appeared in a television ad, which was produced by Radio TV Malacañang and the Philippine Information Agency. In addition, the First Lady agreed to be interviewed for the census. The interview was done in Malacañang by a public school teacher assigned to cover Manila. Likewise, interviews were made for former presidents Corazon C. Aquino and Fidel V. Ramos. All these interviews were captured by the media.

Also through the UPCMCFI, the NSO was able to get a private sponsor, Unilever, to put in resources in exchange for advertising its product in census stickers, which were posted in every housing unit in the country to signify completion of interview. The same worked out for Bangko Sentral ng Pilipinas (BSP) or Central Bank of the Philippines. The product (Domex) logo of Unilever and the BSP slogan with a picture of a bankbook were printed on

the sticker together with the Angat Pinoy logo of the Philippine government's medium-term development plan. This depicted the support and partnership of both the private and government agencies through the Unilever, BSP, and the Angat Pinoy logo.

To monitor census enumeration nationwide, email (computer), fax, land and cellular phones, and telegrams were used by provincial and regional officers to relay problems to regions/central office. Also, the NSO established the “C2K Watch”, an internet-based census monitoring system using maps. The “C2K Watch “ can be browsed by authorized users from NSO through the pages of the NSO web site. The “C2K Watch” is part of an Over-All Quick Count System developed to produce preliminary results from Census 2000. Input to this Over-All Quick Count System emanated from the “Quick Count” forms completed by enumerators that contained population tallies in an enumeration area. Summaries, which were generated from “Quick Count” forms, were encoded by the enumerators using handheld computers. The data contained in the handheld computers were uploaded in the computers in provincial offices that, in turn, transmitted the data to the Central Office via email. At the Central Office, summaries from different provinces were merged and appended to a central database that served as the source of thematic maps for the “C2K Watch” web pages. At the same time, after evaluation, this became the basis of the preliminary results of population count by province and selected cities released in August 21, 2000, in Malacanang.

For the first time in Philippine history of census-taking, the NSO utilized digital processing technology using Intelligent Character Recognition (ICR) in its data processing. NSO acquired more

than 20 scanners distributed in four Data Capture Centers (DCCs), which are located in Manila, Pampanga, Cebu, and Davao. Manila DCC handles questionnaires coming from the field in all the provinces of Southern Tagalog and Bicol and in the cities/municipalities of Metro Manila; Pampanga DCC, those from the Ilocos, Cordillera, Cagayan and Central Luzon Regions; Cebu DCC, those from Central, Eastern and Western Visayas Regions; and Davao DCC, those from the Mindanao regions. In the DCCs, the questionnaires are read by scanners and undergo automatic mark and character recognition and, when needed, manual verification so that the scanned images and database outputs will serve as the basis for the final population counts by *barangay*. Together with the basic data on growth rates, sex ratio, average household size, the final population counts, and distribution will be released in December 2000 and will be proclaimed as official by the President of the Philippines.

Since ICR digital processing is specially sensitive to the quality of the paper, the shade of the ink used in printing the questionnaires and the presence of specific markings on a page, particular attention was given to printing the forms for Census 2000. The scanners used for the image capture of Census 2000 forms reject forms that do not conform to a specific shade of magenta, the drop-out color selected. Also, the printing had been done under strict security to ensure the census materials were not pilfered. The timing of the printing was likewise critical. Census forms and other materials had to be printed early enough to ensure delivery of the materials on time for the training. On the other hand, the printing could not be finished too early, to avoid a long storage period that invites pilferage. It was a massive project for the printing company (a semi-government entity) that was tapped as the exclusive printer for the NSO's Census 2000.

With the application of advanced information and telecommunications technology, preliminary census counts were generated in about only 2 months after the census operations. Similarly, final census counts by *barangay* will be available before the end of this year. Other demographic and socioeconomic characteristics derived from the census will come out in year 2001 (For any inquiries, contact Mr. Tomas P. Africa, Administrator, National Statistics Office, Sta. Mesa, Manila, Philippines, Fax No. (632) 713-7074 and e-mail: T.Africa@mail.census.gov.ph).

POLAND'S 2001 NATIONAL CENSUS

In the 20th century, Poland's national censuses were conducted in 1921, 1931, 1950, 1960, 1970, 1978, and 1988. In 1984 and 1995, so-called "micro-censuses," or large-scale population sample surveys, were conducted. Data were assembled with the use of the traditional method (i.e., register forms were completed in people's homes by properly trained interviewers). During the preparations for the 1988 census, a mixed method of data collection was considered: using both the traditional method and the "self-census," which relied on census forms being filled in only by respondents themselves, without any interviewer participation. Unfortunately, the results of the self-census method were discouraging as only 10 percent of the forms were filled in correctly or with only few errors. The traditional method will be used again in the projected national census, which is to be conducted in April or May 2001. The Polish Parliament has already adopted legislation making the census possible. Basic aims of the census are the following:

1. Collect the demographic, social, and housing data that cannot be collected in any other way.
2. Learn about the changes that have taken place since the last census as far as the basic demographic and social characteristics of the population, households, and housing are concerned.
3. Collect the data necessary for international comparative studies.

The range of the topics included in the census has been decided in accordance with the recommendations of, among others, the UNO, the European Commission, and the Eurostat. In particular, the information relating to the following topics will be collected:

1. Social and demographic topics: geographical structure of the population, the country's demographic picture, structure of households, education levels, cases of disability, people's occupations and sources of incomes, nationalities, languages spoken, religion, homelessness, refugees, etc.

2. Property topics: size, technical equipment and sanitation, heating systems, technical state and age of the buildings, repairs, etc.
3. Other topics: number of children women have, permanent migrations of the people in the years 1989-2000, etc.

A trial census took place of about 50,000 households in April 1999. Its aim was, among others, to check the census methodology, the appropriateness of the questions, people's current addresses, as well as to test processing the data.

After the census proper, a "check census" will be taken to assess the quality of the obtained data. A small sample of households will be drawn according to a sampling plan. Then, once again, a thorough study of the population and households will be made.

Between 130,000 and 200,000 interviewers are to be recruited from among students, the unemployed, OAPs, retired employees, teachers, etc. The estimated cost of the census is US$120 or 140 million. Almost a half of this sum is to be paid to the interviewers.

There are plans for the following three kinds of publications: 1) statistical yearbook, 2) papers referring to Poland's regions, 3) analyses of specific problems.

More detailed information on the projected census may be provided by Ms. Lucyna Przybylska (l.przybylska@stat.gov.pl).

SOUTH AFRICA: CENSUS 2001 IN BRIEF

Legislative Frame

Conducting a population census in South Africa is a statutory requirement. The Statistics Act of 1999 promulgates that Statistics South Africa (Stats SA) conduct a census every 5 years. The first census of the new millenium will be conducted in October 2001. This will be done at a budgeted cost of R 623 million. The strategic goal of Stats SA is to respond to user-defined statistical requirements, by providing small area socio-economic statistics, which are reliable in content and relevant to needs of the users. The Chief Directorate for Demography is the division that is responsible for planning and implementing the various phases of the census. Elaborate plans aimed at making Census 2001 a success are in place and are being rolled out in several areas.

The first census in a democratic dispensation was one conducted in October 1996. Stats SA counted the people in South Africa using a uniform methodology and adjusted the count in accordance with international best practice by a post-enumeration survey (PES). This was the first time in South Africa since 1970 that a complete count was undertaken. In the intervening censuses, large groups of people (e.g., those in the former TBVC states), were not actually counted but estimated by means of demographic projections, while many of those in informal settlements or former homelands were counted by means of aerial photography, or a "sweep census" without prior demarcation.

This first count in an unevenly developed democratic country was a major achievement. Nevertheless, it presented Stats SA with many new challenges. A number of large workshops have been held to learn from previous experiences. In addition, a research-based review of the results, convened by the Statistics Council, has yielded suggested improvements.

Embracing Project Management Ethos

The major difference between Census '96 and Census 2001 will be better programme management, using professional project-managers with their methodology and software. Census '96 was essentially managed as a series of separate phases (i.e., demarcation, enumeration, processing, analysis, and dissemination). Their interdependencies were not always adequately identified. This time, the census is being viewed as an integrated programme consisting of a range of tightly interlinked projects. In addition, there will be alternative service delivery through outsourcing of noncore, corporate services functions in selected respects.

The strategies and logistics of handling these issues are being incorporated in the appropriate plans for Census 2001 as part of the overall project management. For example, the group looking at questionnaire design has suggested rephrasing migration questions. The team looking at

enumeration will develop alternative approaches to access the difficult-to-reach areas. These consist of the very wealthy and commercial farming areas.

Participating to Enhance Best Practice

An advisory committee was formed to assist various aspects of Census '96, including questionnaire design, demarcation, listing, and dissemination. It was made up of people drawn from 35 different organisations, and it had monthly or more frequent meetings. There were also subcommittees handling specific tasks (e.g., questionnaire design and dissemination). A similar committee for Census 2001 has been now formed and has had its first meeting on 22 May 2000.

During Census '96, Stats SA received advice and technical assistance from the United Nations Fund for Population Activities (UNFPA), the Australian Bureau of Statistics (ABS), and Statistics Sweden. In Census 2001, this type of international participation and assistance should be able to continue through our association with Statistics Sweden and probably Statistics Canada. The UNFPA will also continue to provide support through the South African Development Community (SADC) Statistics Committee (SSC).

Stats SA has been working very closely with other SADC countries in relation to identifying core questions to be asked throughout the SADC region in the 2000 – 2001 round of censuses throughout the region. This interaction gives Stats SA an excellent opportunity to learn about census-taking from other African countries.

Stats SA concluded a series of consultative meetings in each of the nine provinces to ask for suggestions for the census questionnaire. A further consultation will involve key local and international users who have made detailed computations on the Census '96 data. In July Stats SA will submit the questionnaire to the Cabinet for comment.

As is common practice in most countries of the world, including those in Africa, there are interministerial committees to ensure an appropriate general work programme for the census, provide political support and legitimacy, and assist in securing logistical and material support from other departments. Such a committee has been constituted in this census and this was not the case in the census of 1996. This committee will receive progress reports from the census advisory committee and the project management team and interact with the Statistics Council in its statutory responsibility for census review.

A multifaceted social marketing publicity campaign has been lodged. The consortium includes public relations, multimedia advertising, and research components. Stats SA, GCIS, and the Consortium have started working out a detailed programme of action for the campaign. Elements of the campaign will define the role to be played by the various levels of political leadership. For instance, the inaugural launches of Census 2001 could be a matter for the individual Premiers in their respective provinces. Protocol for the national launch would likewise need to be established.

Stats SA will be facilitating another workshop in the series, which will deal with Data Processing and Editing, that will be held in Gaborone from 19th June - 7th July 2000.

Canvassing User Needs

Meeting the needs of our domestic users on the other hand, have been an on-going process that started at the time of releasing Census '96 results. As part of consolidating the gains from Census '96 the planning of Census 2001, the questionnaire design team has visited all nine provinces in an attempt to solicit user input into the Census 2001 topics. We further posted advertisements in the print media for that purpose. The end product of that process will be fed into the design of the census questionnaire, which will in turn be submitted to the Census Advisory Committee. This committee, as well as the International Consultants attached to Census 2001, will play a pivotal role in supporting our processes.

The Census Advisory Committee has had its first meeting, were all Census 2001 Project Managers briefed the committee members in Pretoria, on 22 May 2000. The committee has been constituted from several constituencies.

User consultation workshops started on 14 April 2000, and ended on 10 May 2000. Input from various users will continue through email, Internet, telephone, or fax. The users were given important

dates to note and dialogue with users was agreed to be a continuous process.

The user dialogue included the purpose of the census and the comprehensive planned route toward a manageable questionnaire that included the following:

- Fewer topics than in 1996;
- Option of sample surveys to compliment the census;
- More precoded questions;
- Clearer formulation of questions.

So far the responses obtained from the users indicated sentiments toward the census. It will be difficult to strike a balance between user needs and constraints faced by the census. Users view the census as the main source of data and not surveys. The concern about surveys was that although they collect useful information, they are expensive and the involvement of all stakeholders is minimal. Most of them did not know about the sample surveys that are being conducted by Stats SA but were knowledgeable about activities of and publications produced by other departments like Education, Health, and Office of the President and Transport.

Advancing Public-Private Sector Partnerships

An important feature of Census 2001 has been in the deliberate strategy to identify and nurture public/private partnership and intergovernmental collaborations from the start. For instance, the Census 2001 Planning Workshops were facilitated by a private company, which specialises in project management. The clarity and sense of purpose that was injected into the process has been a revelation. As a result of this realisation, Census 2001 sent the project managers to Wits Business School to sharpen their project management skills. This public/private venture is beginning to bear fruit elsewhere in the operational areas of the census programme.

The demarcation phase has always been a very cumbersome and drawn-out process in any classical census-taking. For each enumeration area, from which census data would be collected and disseminated, demarcation officers would need to visit and confirm verbatim the physical features of the enumerator area and its boundaries. For a country the size of South Africa, with about 80,000 enumerator areas to be demarcated, the task could be very daunting and intimidating, even if there was time to spare. But with the advanced techniques of aerial photography and satellite imagery in the marketplace, virtual demarcation can now be done within a matter of months from a desktop computer, situated in any office. Tenders have already been awarded for the acquisition of relevant data sets, as well as expertise in formatting and integrating acquired data, which is aimed at short-circuiting the demarcation process. The preliminary results of field tests from some rural areas, tend to vindicate our virtuous trust of our public/private partnership to deliver quality product on time.

The other partnership that has had a significant impact on our processes is the GIS project called Project Eagle, which was facilitated by the IEC in 1997. Imperfect as it was, that GIS solution has ushered in an unprecedented revolution in the away socioeconomic statistics is being handled. Its power and compatibility with sophisticated dissemination tools was demonstrated during the dissemination phase of Census '96. With this tool at our disposal, Census 2001 has defined the GIS Project as a crosscutting activity, which will supply geo-referenced information of ever aspect of our Census 2001 activities. The system design and hardware and software infrastructure is all being handled by the private sector consortia, that have tendered their services under the aforementioned public/private partnership. The expectation is that some elements of Census 2001 enumeration, logistics and data processing projects will undergo similar exercises.

The downside of these public/private partnerships from the Civil Service point of view, has always been about jobs, or the apparent loss of it to the private sector. But the paroxysm that census or similar projects within government impose on the host organisation has exposed public servants as lacking project management skills. In fact, the basis for their appointments to government positions is nearly always on subject-matter specialisation and/or to become bureaucrats. The challenge in determining the focal point of these partnerships, therefore, has been to build managerial capacity within departments in order for them to be able to manage these complex processes. The other area of benefit is in the fact that the centre of excellence in most high-tech

solutions would invariably lie elsewhere in the private sector. Stats SA is fast catching up with the rest of the industrialised countries in this area.

Gaining International Exposure

In September 1999, two representatives from our division went to England to study the manner in which the British and some European and North American census offices will be processing their Census 2000/1 data. In the British case, the private sector is responsible for the entire process of questionnaire printing. The deliverable to the Office of the National Statistics is clean census data; and all of these occur in a historically conservative environment with relatively hostile public sector unions. The United States of America and the Republic of Ireland have taken similar routes. Even in Zambian, where we sent another delegation in May 2000 during their Census Pilot, has "outsourced" their data processing activities to the country's Examination Council.

Issues for Dissemination

For Census '96, Stats SA was successful in disseminating its Census '96 data because of the early decisions taken to use better technology. Stats SA, however, suffered some setback regarding this technology and had to be creative and innovative to overcome some of the technical limitations of the technology.

Stats SA has been successful in reaching out to the public sector and the private sector, as well as research institutions, and this required investment in infrastructure and training. However, Stats SA has yet to develop strategies for reaching out to the general public, in particular the less literate of the society. Stats SA also has a challenge in relation to pricing its products, especially from the government sector.

Some form of research to evaluate the market has been helpful. This was in terms of positioning and defining what relationship would have to exist between Stats SA and the users. However, the results that relate to prospective numbers that would acquire the data suggest that the research results were pretty conservative, particularly for the public sector.

Technological advancements have been very rapid, and dissemination of the next census could be through fully-fledged web-based techniques with users having full control over the enabling software and data. However, this poses interesting questions on charging and copyright.

SADC Collaboration on the 2000 Round of Censuses

The SADC has established a statistical committee whose aim is to harmonise their data collection methods. In the past 50 years, each member of the SADC has conducted at least three population censuses; eight members are planning such undertaking in the period 2000-2002, and another four are planning mid-term censuses or large-scale demographic household surveys. An assessment seminar on the capability and readiness of the SADC member countries to conduct a population census or large scale demographic household survey revealed a common concern for the inadequate number of technical personnel, equipment, and financial resources for such an undertaking.

The SSC conceived of the first step toward harmonisation through adopting a common approach to the 2000 round of censuses. This will be achieved in the main, by adopting common approaches to data collection, data processing, and subregional training programmes in various activities of census-taking. The successful implementation of the project comes at a time when SADC countries have come to realize the need to harmonize data collection methodologies, terms and concepts, and the dissemination of ensuing statistics as a necessary step toward further integration of economic and social basis of the community. It will also establish mechanism between countries and subregional networking in the field of population/ health statistics. Since census data collection is the largest statistical undertaking in most of the SADC member states, successful implementation of the population census programme will provide the impetus for statistical data harmonization efforts.

NEW PROCEDURES OF THE SPANISH 2001 DEMOGRAPHIC CENSUS

Demographic censuses are one of the major statistical projects periodically to be carried out by the National Statistical Institute.

Since the first modern census in 1768 under Charles III, where the unit of analysis was the individual person, the census-taking process in Spain has never stopped improving.

Despite being a very old statistical tradition, demographic censuses are dynamic and need to be continuously adapted to the country's socioeconomic and even political-administrative situation using the available technology.

An example of this is the Spanish 2001 Demographic Census. In 1996, modifications were introduced in the legislation that regulates the management of the Municipal Population Registers. These cover more than 8,000 municipalities in Spain and report the total number of inhabitants, as well as inhabitants' basic characteristics such as name, place and date of birth, gender, number of identity cards (DNI), nationality, and level of education. This facilitates the relationship between a municipality and its inhabitants. Until 1996, the registers were kept by the municipalities, which supplied additions and deletions without any coordination.

Every 5 years the registers were entirely updated by exhaustively routing the territory and recollecting a new sheet of registration data at each residence. In census years, the census and the update of the register were performed simultaneously.

Today, municipal population register management is computerised and coordinated by the National Statistical Institute. The institute is in charge of checking that all additions and deletions are incorporated and avoiding duplications among municipalities and the register of Spaniards living abroad, which is conducted by the Ministry of Foreign Affairs.

Project implementation has many problems, mostly derived from the slow computerisation of registers in many municipalities, particularly the smaller ones. However, official figures for Spain 1 January 1998, have already been obtained with the new system. By the end of this year, figures will be issued, 1 January 1999, and monthly information on variations will be started. At the same time work is being done to improve the quality of register data. The Census will be the main tool to improve quality.

To make this Census-Municipal Population Register relationship more efficient, census questionnaires will contain personal register data so that each household will receive an envelope with its members' register data, which will only have to be checked and if necessary, corrected.

Since the population census and the Housing and Building Censuses will take place simultaneously, fieldwork regarding personal data will be facilitated by using information from the Cadastre. It will be easier to find the location of the main family households of the census (i.e., those where people are living). On the other hand, the Cadastre will supply information on other types of dwellings (without any registered person) and on premises.

Creating Files for Personalising Questionnaires and Route Books

Creating files to be used to find all the census locations requires integrating information from both administrative registers. Computerized file matching is technically very complicated since file structures, sizes, and coding vary. Furthermore, the files correspond to different reference dates, and, obviously, none of them is without problems.

Bearing in mind that a new way of register management is being implemented, register files should be as accurate as possible. It is, therefore, necessary to take time before setting up personalised files from questionnaires and route books.

Because of these issues the census reference date has been postponed until **the last quarter** of 2001 (census day in 1991 was March 1).

Determining Logical Locations

Simplifying the fieldwork is essential in any census. That is why route books will be personalised through computerised matching the Municipal Population Registers and Cadastre; the addresses of all the census locations in the section on the books will be preprinted.

In order that the route books are as complete as possible and to avoid that the interviewer has to fill in only a few new items, the books will contain preprinted logical locations that are reasonable deductions from the available information. For example, if census locations within a building have been identified through the Municipal Population Register-Cadastre crossing on flats 1º A, 1º C, 2º B, 3º A, 3º B, and 3º C, then flats 1º B, 2º A, and 2º C will be added as possible locations.

Printing Questionnaires

Printing census questionnaires has always been a tricky stage because of the variety of questionnaires needed and the number of copies required for each type. In Spain, it is necessary to issue questionnaires in Spanish, as well as bilingual ones in Spanish and each of the languages of the autonomous communities: Catalan, Basque, Galician, Majorcan, and Valencian. Since the 2001 census questionnaires are going to be personalised, problems become more acute because of the desired quality and the time involved in editing.

Besides the personalization difficulties, there are problems derived from scanning and the optical recognition process, as well as from the questionnaires' minute design.

As to the period for their elaboration, the main inconvenience is that the "enterprises" (publishers) have to print them in the shortest possible time since, as has been pointed out, the printing file should use register data that are nearest to 1 May 2001. Since questionnaire printing takes so long, the dates to which the register data refer should not be after the end of 2000.

Putting the Questionnaires in Envelopes

For the 2001 Demographic Census, four types of questionnaires are going to be used: a register data sheet, a housing questionnaire, a household questionnaire, and an individual questionnaire. They will be delivered to the household in an envelope.

While the Spanish enterprises are able to produce the questionnaires simultaneously for the same household, they are technically not prepared to have them automatically stuffed into envelopes. This operation has to be done manually, which requires supervision to ensure that every household receives all its questionnaires.

Distributing Census Documentation

Because of the personalization and the lengthy printing of the 2001 census documentation, the logistic problem of its distribution will be more complicated.

Consequently, the offices in districts and areas tasked with the census will have to be set up well in advance so that the enterprises may send them the documents as soon as they are ready.

Scanning Questionnaires and Route Books

Most of our neighbouring countries are going to scan their questionnaires. The Spanish National Statistical Institute already envisaged this option for the 1991 census, this year being the starting point for the analysis of available documentation.

Scanning makes it possible to shorten the period to disseminate census results. It reduces expenses since it allows for the destruction of paper questionnaires as soon as they have been digitalized, thus considerably reducing storage costs.

Since census questionnaires are so bulky, top quality, high-speed scanners will be used. The scanners will be rented because it is not possible to reuse them.

Using Character Recognition, Video Recording, and Video Correction

Once the questionnaires are scanned, the digitalized images are recognised. This procedure makes it possible to know the trademarks, bar codes, numerical characters, and even manual literals in the questionnaires.

Video recording will be completed with video correction. Recognized values and their real image will be compared. The computer system is linked to recognition engines that use auxiliary dictionaries that help resolve many recognition problems. The introduction of true values by an operator will only

be necessary when the true image and the recognised one are not the same.

The system includes a series of quality control checks so that when a questionnaire has been read, everything has been done to avoid the need for future access to the original document.

From this point, all readings will be performed automatically. The experience of the 1991 Population Census will be taken advantage of, and as a basic software for the editing, use will be made of the INE created DIA for intrarecords imputation and of Statistics Canada NIM for interrecords imputation.

Filling in Questionnaires by the Internet

Although the Internet is not yet used broadly enough in Spain to use it as an integral part of the census (which would considerably reduce costs), people may fill in the census with the Internet.

The main drawback of using the Internet is the lack of absolutely reliable communications. Therefore, when municipal population register data have to be modified, an electronic signature will be required.

Nevertheless, when census completion implies no changes in the register data, the electronic signature will not be necessary, even though the information will navigate in the Net anonymously and encrypted, should it be intercepted. Every envelope will include an indispensable code for use with the Internet, and the respondent will have to supply information on his or her national identification card, which is not printed on the questionnaire.

This alternative way of filling in the questionnaire hampers fieldwork, which should, therefore, be well coordinated to avoid unnecessary trouble for the population and effort of the collection.

Determining Census Figures

Conventionally, census figures were directly derived from the route books, emphasising the routes and disregarding detection of duplications or false inscriptions, which would have been impossible anyway.

In order to determine the 2001 census figures, modifications introduced by the citizens in their register data will be critically analysed: any entrance should be linked with a withdrawal, except in very specific cases.

Consequently, information supplied by the citizens will be checked against the original data made on people on the questionnaires and some auxiliary facts (the accuracy of their DNI's, their names on the electoral roll, their potential duplications after verification of the Municipal Population Register, their names in deaths and births files, coincidence with returned census roll cards, etc.)

In cases of ambiguity, all this information will be used to determine whether they are to be included in the census figures or not. Summarising, ambiguous cases are those people collected during the census and who are missing in the initial files, as well as those who did appear in initial files but have not been found in the census route. However, an analysis will also be made of those initial inscriptions for which we have questions suggested by auxiliary information.

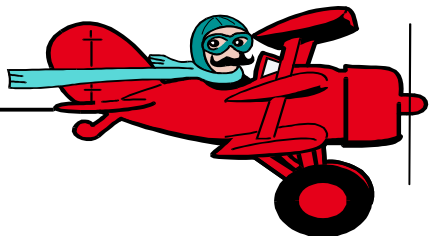
Disseminating Census Results

The entire census approach is meant to shorten the periods to the dissemination of results. As stated in the Census Project, the dissemination stages are:

- By **July 2002**, census population figures will be disseminated down to the municipality. From then onward, structures will be disseminated by gender and age.
- By the end of 2002, the remaining register variables will be disseminated (place of birth, nationality, residence in the year before, etc.).
- By the end of 2003, the census variables will be disseminated according to the standard of publications.

Along with dissemination by conventional procedures (paper and electronic publications), the Internet will be widely used. Apart from the established plan, requests for tailored information will be satisfied, which is also possible by the Internet. In order to proceed more quickly, census information will be stored in a data warehouse.

Welcome New Members!



We are very pleased to welcome the following new members.

Albania

GENT HASHORVA

FATMIR MEMA

Algeria

KHALIDA BOUDALI

Argentina

GONZALO PABLO DOMINGO MARI

Austria

EWALD KUTZENBERGER

Belgium

ABDELGHAFOUR AYADI

Canada

FRANCINE HARDY

SYLVIE MICHAUD

China

HING-WANG FUNG

France

JOSE BARDAJI

ANNE GAYET

ELIZABETH OMOLUABI-SIEGERT

RICHARD STOECKEL

LAURENT TOULEMON

Gambia

OLAJIDE EZEKIEL OKUNOLA

Greece

MARIA KARAKLIOUMI

Haiti

MARIE SUZIE JOHANNE BEAUCHAMPS

PIERRE JACQUES VIL

Hungary

ERZSEBET EILER

India

KRISHNA MOHAN

Italy

NAMAN KEITA

Ivory Coast

RAILOVY

MAHENINA RANAIVO

Lebanon

NADER KEYROUZ

Madagascar

ARSENE RAVELO

Mauritius

H. BUNDHOO

Nepal

SHYAM BHAKTA UPADHYAYA

Niger

ABOUBACAR GADO

SANI OUMAROU

Nigeria

MARUF KOLAPO USMAN

Philippines

OLIVIA PAGULAYAN

Poland

JACEK S. KOTOWSKI

Russia

OLGA MIRHAILOVNA MOUNATCHEVA

Senegal

MADY DANSOKHO

BABAKAR FALL

ALIOUNE NDONG

OUMY SARR

Slovenia

EVA BELAK

KATARINA KOSMELJ

Sri Lanka

SELLIAH SIVARAJASINGHAM

Switzerland

A. SYLVESTER YOUNG

Tanzania

ALBINA CHUWA

Togo

AGBENOZAN JEAN TETTEKPOE

Tunisia

CHEDLI BEJI

Turkey

A. SINAN TURKYILMAZ

Uganda

YOVANI ADULAMU MOSES LUBAALE

Ukrania

OLERSANDR GLADUN

VOLODYMYZ SARIOGLO

United Kingdom

ROBERT J. BUTCHER

ANDREW MACCULLOCH

PARMEET SINGH

United States

PROMOD CHANDHOK

SUBIR GHOSH

DENNIS HANSEMAN

JASON HSIA

MADELINE MCDEELEY

ROBERT F. RAUTENBERG

Zambia

PRISCILLA CHILOMBO MUSOLE

Institutional Members

Mauritius

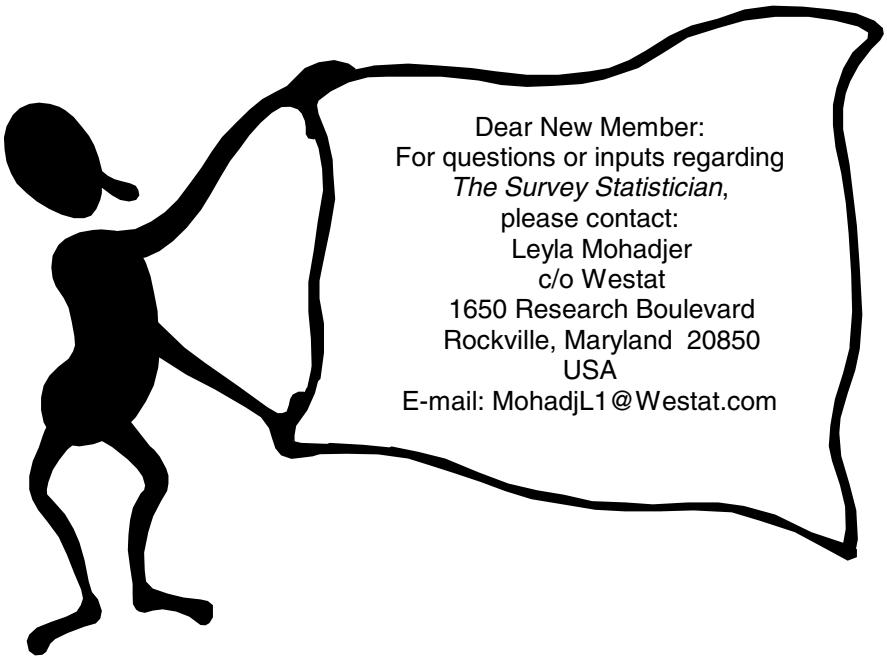
CENTRAL STATISTICAL OFFICE

Netherlands

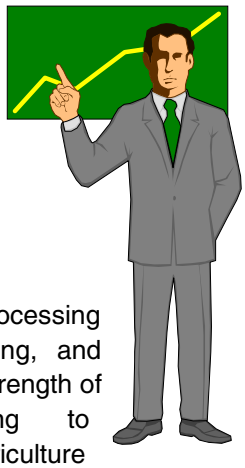
CENTRAL BUREAU OF STATISTICS

Tanzania

BUREAU OF STATISTICS



CONFΣRΣNCΣ RΣPORTS



ICES-II

Dan Kasprzyk

The International Conference on Establishment Surveys (ICES-II) was held in Buffalo, New York, from June 17 through June 21. A major strength of the conference was the strong international presence, both on the program and in attendance. About 470 people from 35 countries took part in the conference. Among those attending were groups from Italy, Sweden, and England, as well as participants from Indonesia, Surinam, Ethiopia, Finland, Slovenia, China, and many other countries. The conference enjoyed the support of the International Association of Survey Statisticians, the ASA Sections on Survey Research Methods and Government Statistics, and the Statistical Society of Canada, as well as many government statistical agencies and private survey organizations.

John Kovar, Chair of the Conference Committee, opened the conference by welcoming everyone to Buffalo, while Susan Linacre from the Office for National Statistics in the United Kingdom kicked off the conference with the keynote address. Susan looked back at the problems facing establishment surveys at the time of the first ICES, observed how they have gradually evolved in the intervening years, and discussed important issues to be addressed in the coming years.

The program included 33 invited sessions, 26 contributed sessions, 28 software demonstrations, and 3 short courses. The smooth operation could be attributed to the fine work of staff from the U.S. Census Bureau, Statistics Canada, the American Statistical Association, the National Agricultural Statistics Service, and the Adam's Mark Hotel, all under John Kovar's direction.

Many other topics were the focus of sessions and software demonstrations, including technological improvements for collecting and disseminating establishment data, classification, school sampling frames, the use of administrative records, generalized estimation and processing systems, coordinated sampling, and time series. Because of the strength of the presentations relating to agriculture, the Food and Agriculture Organization of the United Nations has sought permission to publish them in a separate volume.

Around the beginning of 2001, invited papers from the conference will be available in a hard-covered proceedings book. Invited *and* contributed papers will be included on a CD-ROM. For purchase information, contact Statistics Canada at kovar@statcan.ca. The book and the CD-ROM will be sent to all conference attendees.

Colloque francophone sur les sondages Brussels-Lille, June 21 to 24, 2000

Benoît Riandey

One year after the colloquium on "Surveys and Information Systems" in Abidjan, the Colloque francophone sur les sondages was organized in Brussels, Belgium, by Jean-Jacques Droesbeke. The colloquium was held in Brussels and Lille from June 21 to 24, 2000, with one hundred and fifty participants from ten or so countries, including a large Canadian contingent. Ludovic Lebart (France) was Chairman of the Science Committee while Carl Särndal (Canada-Finland) was Vice-Chairman.

A "Surveys and Politics" Day was held in partnership with the CRAPS and the Political Science Division of l'Université Libre de Bruxelles, in Lille, France, near Brussels, on June 21. The main colloquium took place at l'Université Libre de Bruxelles. Nine guest lecturers and forty-five unsolicited presentations provided a range of information regarding survey and estimation theory, the history of surveys, the census, collection methodology, special surveys on audience rating,

health, violence, poverty, heritage, employment and business surveys. L'École de commerce Solvay hosted a "Surveys and Businesses" Day on June 24 in Brussels.

The colloquium proceedings will be published as soon as possible, following those of the Rennes colloquium: *Enquêtes et sondages : méthodes, modèles, applications, nouvelles approches*, under the direction of Gildas Brossier and Anne-Marie Dussaix, 1999, Dunod. The next colloquium will be held on October 17 and 18, 2002, in Grenoble, France, and will include a satellite day on the subject of political surveys.

Announcements

Workshop on Small Domain Estimation in Labour Force Surveys

The workshop is scheduled to take place April 3-6, 2001 in Libourne, France. It will be hosted by Cefil, the Insee facilities in Libourne where the Secretariat of IASS is located. The 3-day workshop will give experts on labor force surveys the opportunity to discuss issues related to small domain estimation.

Estimates for small areas will be considered, as well as estimates for small population subgroups. The detailed program is being put together under the guidance of Farhad Mehran (mehran@ilo.org), Vice President of IASS.

Preliminary programme:

1. Labour force survey - based estimates for small areas

   - Role of administrative records
   - Estimation methods (the notion of "borrowing strength")
   - Assessment of bias
   - Uses and risks in allocation of government funds
   - Presentation of results to the public
   - Special issues in countries with low coverage of unemployment insurance systems or low use of employment offices as a method of job search.

2. Labour force survey - based estimates for small population subgroups

   - Relevance of small area estimation techniques
   - Foreigners and immigrants
   - Low-wage workers
   - Disabled workers
   - Child labour
   - Home workers
   - Street vendors

The programme will be finalised on the basis of the submission of the contributed papers and other proposals.

IASS 2001 Program

The June edition of the Survey Statistician included a report on the IASS program for the 2001 ISI meetings. An updated list of the currently planned invited paper meetings sponsored by the IASS program is as follows:

IASS organisers

The Role of Survey Sampling in the 21st Century.
Organiser: John Cornish

Combining Data from Different Sources.
Organiser: Tim Holt

Misclassification as Response Error in Surveys.
Organiser: Cathy Dippo

Internet and Innovative Data Collection.
Organiser: Warren Mitofsky

Disclosure Control and Data Access.
Organiser: Luigi Biggeri

Multiple Frame Surveys (with Korea/ISI).
Organiser: Alvaro Gonzalez-Villalobos

Linked Employer-Employee Data.
Organiser: Cynthia Clark

Multilevel Models for Survey Design and Analysis.
Organiser: Chris Skinner

Price/Production Indices.
Organiser: Dennis Trewin

Edit and Imputation Techniques.
Organiser: John Kovar

Joint Sessions being organised by IAOS

Quality Programs in Statistical Agencies.
Organiser: Gordon Brackstone

Measurement of E-commerce.
Organiser: Paul Cheung

For information on the ISI meetings, readers should check out the web site
at http://www.nso.go.kr/isi2001/.

IASS Short Courses at the 2001 ISI Meetings

The International Association of Survey Statisticians has developed a program of short courses to take place before the 53rd Session of the International Statistical Institute (ISI), which will be held in Seoul, Korea, August 22-29, 2001. The courses, led by international experts in their fields, will be useful to practitioners, researchers, and students in statistics and survey methodology. All courses will be presented in English, and participants should have the ability to work in this language. The courses will be held in Seoul at the location where the ISI sessions will take place.

The following courses are scheduled to take place just prior to the ISI meetings — during the period August 18, 2001, through August 22, 2001.

Course A: Survey Sampling

August 18, August 19, August 20 (morning), 2001
Presented by:
Graham Kalton, Westat, Rockville, Maryland, U.S.A.
Colm O'Muircheartaigh, National Opinion Research Center and Harris School of Public Policy Studies, University of Chicago, Chicago, Illinois, U.S.A.

This workshop is designed for participants who have practical survey experience but who do not have specialist expertise in survey sampling. It will be of particular benefit to those involved in survey sampling in developing countries and countries in transition.

The workshop will focus on practical applications rather than sampling theory. Relevant mathematical formulas will be presented and discussed, but no mathematical results will be derived. The workshop will start from a review of basic principles and build up to sample designs for national surveys. Simple exercises will be given at various stages to reinforce important aspects of sample design.

Topics to be covered include the following: sampling frames and frame problems; various sample designs — simple random sampling, systematic sampling, stratification, cluster and multi-stage sampling, and probability proportional to size sampling; weighting for unequal selection probabilities, nonresponse adjustments, and poststratification adjustments; and an introduction to variance estimation with complex sample designs.

Course B: Variance Estimation in Complex Surveys

August 20 (afternoon), August 21, August 22 (morning), 2001
Presented by:
Wayne Fuller, Iowa State University, Ames, Iowa, U.S.A.
Kirk Wolter, National Opinion Research Center, Chicago, Illinois, U.S.A.
F. Jay Breidt, Iowa State University, Ames, Iowa, U.S.A.
Jae-Kwang Kim, Westat, Rockville, Maryland, U.S.A.

This course will provide training in variance estimation in complex surveys for survey statisticians, especially for those from developing countries. The course will cover methods of estimating variances for statistics such as means, proportions, ratios, regression coefficients, and statistics arising in the analysis of two-way contingency tables.

Both linearization and replication methods will be discussed. The use of computer software for computing variances of statistics from complex sample designs will be demonstrated, and instruction will be given in practical applications. About one-half of the course will be devoted to methodological considerations and one-half to implementation and computation.

Course C: Introduction to Small Area Estimation

August 20 (afternoon), August 21, August 22 (morning), 2001
Presented by:
Jon N. K. Rao, Ottawa, Canada

Sample survey data can be used to derive reliable estimates for large areas, but sample sizes in small areas or domains are seldom large enough for direct estimates to provide adequate precision for such areas. This makes it necessary to use indirect estimates that borrow strength from related areas. This short course will provide an overview of indirect estimation methods, traditional as well as model-based. Traditional methods include synthetic, composite, SPREE, and sample size dependent methods. Model-based methods use random effects models that relate a characteristic of interest to auxiliary variables with known population characteristics. Such methods include empirical best linear unbiased prediction (EBLUP). Empirical Bayes (EB) and Hierarchical Bayes (HB) methods under two types of models: area level and unit level. Methods for measuring the variability of the estimates as well as model validation will be discussed. Extensions to binary data and time series data will be covered. Several recent applications will be discussed. Methods will be compared on a simulated population resembling a real population.

Course D: Nonsampling Error Research

August 20 (afternoon), August 21, August 22 (morning), 2001
Presented by:
Clyde Tucker, U.S. Bureau of Labor Statistics, Washington, D.C., U.S.A.

Nonsampling error is a persistent and important source of error in sample surveys. This course will provide a brief overview of nonsampling error in surveys, both household and establishment surveys. The philosophical, theoretical, and practical aspects of nonsampling error research will be discussed with special emphasis on recent work conducted by the U.S. Bureau of Labor Statistics to measure various nonsampling error effects.

Survey researchers are able to obtain a good estimate of the variance component of mean squared error. Unfortunately, estimating the bias (nonsampling error) component is more difficult and usually not undertaken. Because the magnitude of the nonsampling error component is usually larger than the sampling error in estimates from large sample surveys, it is important to learn more about this component of error. Those interested in understanding the extent of nonsampling error in their surveys should find this course useful. There are no prerequisites, but some knowledge of behavioral science applied in survey methods would be helpful.

Course E: Editing and Imputation of Survey Data

August 20 (afternoon), August 21, August 22 (morning), 2001
Presented by:
John Kovar, Statistics Canada, Ottawa, Canada
Eric Rancourt, Statistics Canada, Ottawa, Canada

Surveys and censuses conducted by national statistical agencies, research institutes, and other survey organizations incur various degrees of nonresponse even under ideal conditions. To alleviate the problems caused by nonresponse, editing and imputation methods are usually applied. Because nonrespondents may not behave in the same manner as respondents in terms of characteristics of interest, special care must be taken to avoid potential biases that may result from inappropriate treatment of missing or inconsistent data. Since the process of editing and imputation is time and resource intensive, care must be exercised in controlling the efficiency as well as the effectiveness of the methods.

The aim of this short course is to introduce the students to methods of prevention, detection, and treatment of nonresponse as well as the evaluation of such methods. Emphasis will be placed on editing and imputation. Types of edits and their application will be discussed. Then, imputation methods will be described along with their advantages and disadvantages. Existing edit and imputation software will be compared. Evaluation strategies will be discussed and the impact of methods on the survey output will be outlined. During the course, various examples will be given to illustrate the material presented.

Morning session classes are scheduled from 8:30 a.m. to 12 p.m.; afternoon session classes are

scheduled from 1:30 p.m. to 5 p.m. The meetings of the International Statistical Institute begin on the afternoon of August 22. The short courses will be completed by lunch on August 22 so that all can attend the ISI opening session and social. All courses will take place at the site of the ISI meetings.

A "welcome social" is planned for all students and presenters on Monday evening August 20, 2001.

Funding assistance to provide travel support to statisticians from developing countries and countries in transition is being sought.

For additional information, contact:
IASS Scientific Secretary, Dan Kasprzyk
(daniel_kasprzyk@ed.gov).

XVIIIth International Symposium
on Methodological Issues
Achieving Data Quality in a Statistical Agency:
A Methodological Perspective
October 17-19, 2001
Statistics Canada

With its eighteenth annual symposium, Statistics Canada continues its successful series of conferences on methodological issues, attracting renowned statisticians, researchers, academics, and data analysts and others interested in meeting the challenges of a statistical agency. Symposium 2001 will feature both invited and contributed sessions, and will provide an ideal forum for exchanging your experiences and knowledge of methods to achieve data quality. Papers of this conference will be published in the proceedings of the Symposium. The symposium will be held at the Palais des congrès in Hull, Quebec, just minutes away from downtown Ottawa. For more information, please contact Simon Cheung, Symposium 2001 Co-ordinator, 16th Floor, R.H. Coats Building, Statistics Canada, Ottawa, Canada, K1A 0T6. E-mail: symposium2001@statcan.ca Telephone: (613) 951-1482 Fax: (613) 951-3100.

Visit our Web site:
www.statcan.ca/english/services/smnrs.htm.

The following technical paper is available on line:

Technical Paper 63: Current Population Survey (CPS) – Design and Methodology was published and issued March 2000. The CPS is the primary source of labor force statistics for the U.S. population. It is sponsored jointly by the U.S. Census Bureau and the U.S. Bureau of Labor Statistics.

This technical paper describes the many changes made to the survey since the publication of the previous technical documentation (*Technical Paper 40*) over 2 decades ago. While the basic approach to collecting labor force and other data through the CPS has remained intact, many changes have been made to the survey, including the improvement of numerous questions and the computerization of the survey instrument, both introduced January 1994.

The document describes the CPS design and methodology as of December 1995. Appendices cover updates that have been made to the survey since then. The document is available via the Internet at the CPS website at http://www.bls.census.gov/cps. Future changes to the survey will be documented at this site.

The booklet of **the Proceedings of the IASS topics in Helsinki** has been mailed to all the IASS members. If you are an IASS member and have not received a copy, please contact:

Ms. Claude Olivier
IASS secretary
INSEE/CEFIL
3, rue de la Cité
33500 Libourne
France
Email: claude.olivier@insee.fr
Fax: + 33 5 57 55 56 20.

International Flavors

Spring as the Statistician Sees It

Words by T.H. Tyler
Drawings adapted from Hazel Luthena Beach
(poem reprinted from THE MATHEMATICS TEACHER Vol. 34, No. 4; 1941!)

Ah Spring, sweet quadrant of the year!
Kind welcomed variation
That brings from Winter's zero rank
A mild attenuation.

We leptokurtics exercise
And strive for restitution
Of platykurtic curvature
Or normal distribution.

Tall ogives bloom in half our yards*
With colors gay and nifty
(The semi-interquartile range
Approximates point fifty.)

Huge quartiles chortle o'er the range
And 'long with each comes trooping
Her small percentiles, happy in
Their homogeneous grouping.

The trees their long abscissas stretch
To left and right of zero
And form a shaded area for
Our heroine and hero,

The horizon's rough sketch he scans
Till he her image spies
And there observes a set of curves
More pleasing to the eyes.

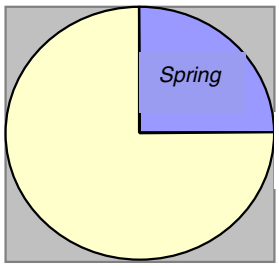
They tread an asymmetric path

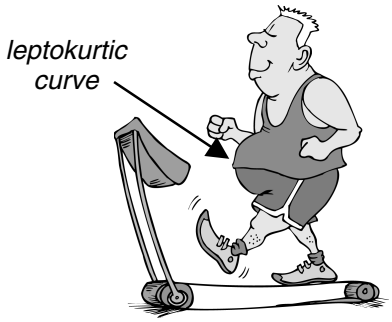
And now and then glimpse heaven
As kisses hourly they exchange
(A mean of two-point-seven.)

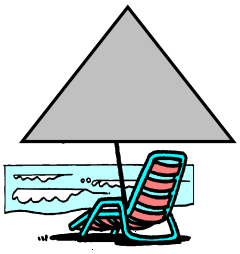
The dainty gamma nods it head
Likewise the flowering beta
As he to her doth soft relate
Ungrouped and spurious data.

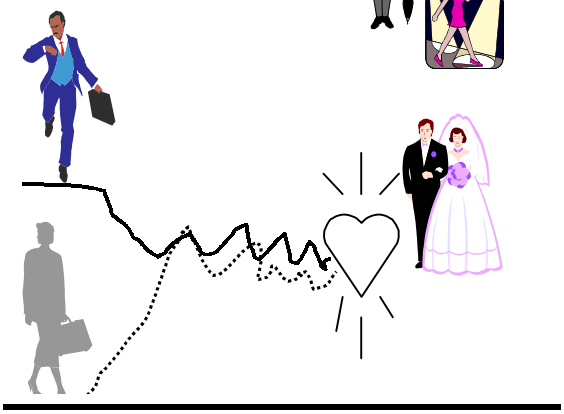
And what results accrue from this?
It leads to that condition
In which the plots of both their lives
Find superimposition.

Thus spring doth bring uncounted charms,
Of infinite summation;
And yet they almost all fall in
Three sigmas variation.









** Now all you sharks don't check too close*
This springtime meditation
Wherein poor meter and poor math
Stand high in correlation;

For if this problem seems confused
Disorganized and muddy,
That simply goes to prove its worth
And need of further study.

The following provides a recipe for success in survey organizations. Follow the steps below when managing projects, training statisticians, interviewers, and other employees, and the outcome should be a job exceptionally well done…

Pride

by Edward C. Bryant

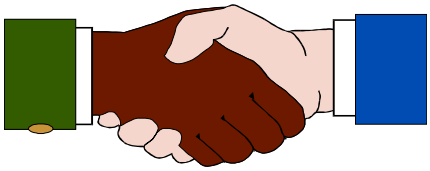
Pride is the first of the seven deadly sins, the others being covetousness, lust, anger, gluttony, envy, and sloth. Presumably, they are "deadly" because they are fatal to spiritual progress, however that may be interpreted. I think pride has got a bum rap. There is both good pride and bad pride, and Webster gives a long list of the kinds of pride considered bad. In fairness, Webster also says "… pride may be commendatory in indicating a justified self esteem, proper self-respect, [and so on]." It is that kind of pride that I am writing about here.

Dr. W. Edwards Deming frequently wrote and spoke about pride in one's work as being one of the principal factors in the success of an enterprise. He pointed out that it is management's responsibility to create a working environment and a management attitude toward employees that permit them to take pride in the quality of their work. This responsibility includes "participative management" in which the manager works with the employees in solving problems rather than directing them to do their work in a given way. This approach helps to break the barriers between departments and between employees and their supervisors. Most importantly, it helps to eliminate the fear of supervision which prevents employees from communicating freely and honestly with management. Deming's philosophy of management has had a tremendous impact on management world wide, an effect that has improved the quality of products, a decrease in management overhead, an increase in company profits, and happier work forces, internationally.

Participative management leads to choosing the best way of doing a job and it results from the interaction of workers and management at all levels. This is referred to as "best current practices," which implies that, no matter how good the methods are, they can be improved, and the operational goal is to improve continuously. Training sessions that reflect many years of experience and interaction between workers and supervisors are more meaningful than sessions based on a rigid set of rules. Employees are likely to be interested in the job when they realize that (1) the job is important, and (2) they have a hand in establishing the rules.

A successful company wants people to be proud of their work and, hopefully, to get a pat on the back and praise for a job well done. That can't always be done, but how the person feels about his or her work is more important than recognition for it. The person himself or herself knows whether he or she has done a good job, and there is nothing to be ashamed of for feeling good about it. That may be pride, but it is a wholesome kind of pride.

It is important to note that pride in the job isn't something that can be directed from above. It is a very personal thing. Generally the employee knows that he or she has done a good job. He or she doesn't have to be told that. This is not to say that recognition for good work is unnecessary—it really is very important. But the emotional boost to one's pride for work well done is the important thing. That kind of pride leads to high quality work, happy workers, a successful project, and an exceptional company.



In Other Journals

Survey Methodology

A Journal Published by Statistics Canada

Contents Volume 26, Number 1, June 2000

Journal of Official Statistics

An International Review Published by Statistics Sweden
http://www.jos.nu/

JOS is a scholarly quarterly that specializes in statistical methodology and applications. Survey methodology and other issues pertinent to the production of statistics at national offices and other statistical organizations are emphasized. All manuscripts are rigorously reviewed by independent referees and members of the Editorial Board.

Contents Volume 16, Number 2, 2000

Contents Volume 16, Number 3, 2000

Journal of Official Statistics

An International Review Published by Statistics Sweden
http://www.jos.nu/

Contents Volume 16, Number 3, 2000 (cont.)

All inquiries about submissions and subscriptions should be directed to the Chief Editor:
Lars Lyberg, R&D Department, Statistics Sweden, Box 24 300, S-104 51 Stockholm, Sweden.

Statistics in Transition

Journal of the Polish Statistical Association

Contents Volume 4, Number 5, June 2000

Statistics in Transition

Journal of the Polish Statistical Association

Contents Volume 4, Number 5, June 2000 (cont.)

Amendment

For subscribing to the journal or sending a paper, please contact Prof. Jan Kordos;
e-mail: j.kordos@stat.gov.pl. Both latest issues of Statistics in Transition may be found at the following web site address: http://www.stat.gov.pl/english/transition.htm.

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

MEMBERSHIP APPLICATION FORM

Please complete the following information if you wish to enroll as a member of the International Association of Survey Statisticians (IASS).

Name: Mr./Mrs./Miss/Ms. _____ First name: _____

Date and place of birth: _____

Citizenship: _____

Occupation *(if retired, please indicate)*: _____

Title (professor, doctor): _____

E-mail address *(please just indicate one)*: _____

May we list your e-mail address on the IASS web site?

Yes ☐ No ☐

Home address

Street: _____

City: _____

State/Province: _____ Zip/Postal code: _____

Country: _____

Telephone number: _____

Fax number: _____

Business address

Company: _____

Street: _____

City: _____

State/Province: _____ Zip/Postal code: _____

Country: _____

Telephone number and extension: _____

Fax number: _____

Please specify address to which your IASS correspondence should be sent:

Home ☐ Business ☐

Are you a member of the ISI?

Yes ☐ No ☐

The official language of your IASS correspondence should be:

English ☐ French ☐

Which field(s) of survey work are you especially interested in?

1. Theory of surveys ☐
2. Survey design and selection ☐
3. Collection of data ☐
4. Data processing ☐
5. Quality of data: errors of measurement, sampling errors, nonsampling errors ☐
6. Analysis of survey data ☐
7. Other *(please specify)*: _____ ☐

Which field(s) of application are you especially interested in?

1. Demographic statistics ☐
2. Housing statistics ☐
3. Industrial statistics ☐
4. Statistics of commerce and service ☐
5. Agricultural statistics ☐
6. Socio-economic studies of households ☐
7. Market research and opinion surveys ☐
8. Education ☐
9. Environment ☐
10. Health ☐
11. Other *(please specify)*: _____ ☐

Please be aware that the membership dues have been fixed at 130 French Francs (US$20, or the equivalent in any other convertible currency at the rate of exchange on the date of payment) and that these dues have been reduced to 65 French Francs (US$10) for nationals of developing countries.

Method of payment:

☐ Check payable to the International Association of Survey Statisticians
Mail directly to:
Secrétariat de l'AISE/IASS
c/o INSEE-CEFIL
Att. Mme Claude Olivier
3, rue de la Cité
33500 Libourne
France
Tel: +33 5 57 55 56 17
Fax: +33 5 57 55 56 20
E-mail: claude.olivier@insee.fr

☐ Bank transfer directly created to the IASS bank account with the Banque Nationale de Paris, in Paris (Bank Code 30004; Agency Code 00874; Account No. 00 45 86 32 Key 88) *(Mention the name and address of the establishment issuing the order and the number of the order.)*

☐ Bank credit card: Visa, Mastercard, Eurocard, American Express *(please circle)*
Expiration date: _____
Card number: _____

☐ UNESCO Coupon

Signature: _____ Date: _____

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

CHANGE OF ADDRESS FORM

If your home or business address has changed, please copy, complete, and mail this form to:

IASS Secretariat
c/o INSEE-CEFIL
Att. Ms. Claude Olivier
3, rue de la Cité
33500 Libourne – France

Name: Mr./Mrs./Miss/Ms. _____ First name: _____

E-mail address *(please just indicate one)*: _____

May we list your e-mail address on the IASS web site?

Yes ☐ No ☐

Home address

Street: _____

City: _____

State/Province: _____ Zip/Postal code: _____

Country: _____

Telephone number: _____

Fax number: _____

Business address

Company: _____

Street: _____

City: _____

State/Province: _____ Zip/Postal code: _____

Country: _____

Telephone number and extension: _____

Fax number: _____

Please specify address to which your IASS correspondence should be sent:

Home ☐ Business ☐

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

Institutional Members

INTERNATIONALE ASSOCIATION D'ENQUETES DES STATISTICIENS

4 international organizations

AFRISTAT
CICRED
EUROSTAT
UNITED NATIONS STATISTICAL DIVISION

30 bureaus of statistics

ARGENTINA - INSTITUTO NACIONAL DE ESTADISTICA Y CENSOS - INDEC
AUSTRALIA - AUSTRALIAN BUREAU OF STATISTICS
BELGIUM - INSTITUT NATIONAL DE STATISTIQUE
BRAZIL - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATISTICA - IBGE
CANADA - STATISTICS CANADA
CHINA - GOVERNO DE MACAU
CÔTE D'IVOIRE - INSTITUT NATIONAL DE LA STATISTIQUE
CZECH REPUBLIC – CZECH STATISTICAL OFFICE
DENMARK - DANMARKS STATISTICS
FINLAND STATISTICS
FRANCE - INSTITUT NATIONAL DE STATISTIQUE ET D'ÉTUDES ÉCONOMIQUES - INSEE
GAMBIA - CENTRAL STATISTICS DEPARTMENT (GAMBIA)
GERMANY - STATISTICHE BUNDESAMT
GREECE - NATIONAL STATISTICAL SERVICE OF GREECE
IRAN - STATISTICAL CENTER OF IRAN
ITALY - INSTITUTO CENTRALE DI STATISTICA - ISTAT
MEXICO - INSTITUTO NACIONAL DE ESTADISTICA, GEOGRAFIA E INFORMATICA - INEGI
NETHERLANDS – CENTRAL BUREAU OF STATISTICS
NEW ZEALAND - STATISTICS NEW ZEALAND
NIGERIA - FEDERAL OFFICE OF STATISTICS
NORWAY - CENTRAL BUREAU OF STATISTICS
PORTUGAL - INSTITUTO NACIONAL DE ESTATISTICA - INE
REPUBLIC OF KOREA - NATIONAL STATISTICAL OFFICE - NSO
SWEDEN - STATISTICS SWEDEN
SWITZERLAND - OFFICE FÉDÉRAL DE STATISTIQUE
TANZANIA – BUREAU OF STATISTICS
UNITED KINGDOM - OFFICE FOR NATIONAL STATISTICS
USA - BUREAU OF THE CENSUS
USA - DEPARTMENT OF EDUCATION
USA - DEPARTMENT OF HEALTH AND HUMAN SERVICES

6 universities, research centers, private firms of statistics

ARGENTINA - UNIVERSITAD NACIONAL DE TRES DE FEBRERO
DENMARK - SFI
EUROPE - A.C. NIELSEN MANAGEMENT SERVICE
FRANCE - INSTITUT NATIONAL D'ÉTUDES DÉMOGRAPHIQUES - INED
USA - RESEARCH TRIANGLE INSTITUTE
USA - WESTAT