Editors
Leyla Mohadjer
Jairo Arrow

Section Editors
John Kovar — Country Reports
James Lepkowski — Software Review

Production Editors
Therese Kmieciak

Circulation
Claude Olivier
Anne-Marie Vespa-Leyder

The Survey Statistician is published twice a year in English and French 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 LeylaMohadjer@Westat.com or mailed to:

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

In This Issue

No. 47, January 2003

Letter from the President

When this issue of *The Survey Statistician* appears, we will be greeting the year 2003. My best wishes to all IASS members and your families. The year 2003 will be a convention year: Berlin will host you from August 13 to 20 for the 54th Session of the ISI. I hope that many of you will be able to make your way to the German capital to show the vitality of our association.

Until we meet in a few months, I would like, as usual, to give you a little information on developments in the IASS.

Administrative Matters

Our Executive Director, Christophe Lefranc, is leaving us after only two years in his position. Having received an assignment that takes him away not only from the Secretariat but also from France and indeed Europe, he will no longer be able to remain in his position. On learning of his departure, my predecessor Kirk Wolter, who worked with him during the second year of Kirk's presidency, wrote, "I must say that Christophe did a fabulous job. We will miss him." I join Kirk in his appreciation. IASS will miss him; I will miss him personally as president, but am glad that he could be replaced under the best of conditions. Alain Charraud, who served as Executive Director, between 1993 and 1995, agreed, with the consent of INSEE, to take up the torch again. I welcome him with the certainty that the management of the Association will be in good hands. He may be contacted at the following address: alain.charraud@insee.fr.

IASS Program Committee for the Sydney Session (2005)

Pedro Silva has assembled his Committee, which has 26 members. They are working on defining the themes of the meetings organized by the IASS, alone or in association with the ISI or another section. They are to send the results of their efforts to the ISI Program Co-ordinating Committee, chaired by Professor Morgenthaler, by mid-June 2003. During the Berlin Session, in August, the Co-ordinating Committee will decide on the final list of themes selected. Elsewhere in this issue is a call for suggestions, issued by Pedro Silva.

Nominating Committee

I had already announced in *The Survey Statistician*, issue no. 45, December 2001 that O. O. Ajayi would chair the Nominating Committee. I am now able to provide you with the list of its members: Jacques Charmes (France), Daniel Kasprzyk (USA), Janis Lapins (Latvia), Diana Masone (USA), Alice Odounfa (Ivory Coast), Claudio Quintano (Italy), Awa Thiongane (Senegal).

Publications

The proceedings of the Seoul Session have been prepared, as is the tradition, by the person who was Chair of the IASS Program Committee for this session, David Binder, who is now Vice-President of our Association. Eric Rancourt assisted him in this task, and I thank them both for this major effort. The proceedings contain the full text of the presentations made in the meetings organized by the IASS, either alone or in association with the ISI or another section. The collection was printed and distributed by the Australian Bureau of Statistics (ABS). I wish to thank everyone who contributed to this undertaking within the ABS and especially Dennis Trewin, who is also President of the ISI.

Work is well under way on the book commemorating Leslie Kish, which contains a selection of his most memorable articles. It should come out in early 2003.

Joint Conference of the IAOS and the IASS in 2004

In the last issue of *The Survey Statistician*, I mentioned the joint conference of the two associations that is to take place in 2004 in Abidjan, Côte d'Ivoire. Things have moved along since then.

In the first place, we worked to establish the Program Committee, which is chaired by Alain Azouvi, formerly of INSEE (France), now retired but still active, notably as ISI representative to the PARIS21 consortium. The following persons have agreed to serve on the Committee: Carmen Feijo (Brazil), Naman Keita (FAO), Zarylbek Kudabaev (Kirgizistan), Achille Lemmi (Italy), Koffi Nguessan (Côte d'Ivoire) and André Portella (African Development Bank). We hope that a few others will be able to join them.

Also, in conjunction with the Conference organized by the IAOS to take place in London at the end of August on the theme of "Official Statistics and the New Economy," a meeting was held that brought together Paul Cheung, President of the IAOS, Luigi Biggeri, President-elect of the IASS (who attended in my place), Alain Azouvi, Marcel Van den Broecke, Director of the ISI Permanent Office, and Jean-Louis Bodin, Past President of the ISI and advisor to the local organizing committee. The Director of Statistics of Côte d'Ivoire, Meleu Mathieu, was unfortunately unable to attend. This meeting launched efforts to develop the program of the Conference, along the general lines of "Poverty, Social Exclusion and Development."

Those who are interested in any way by the theme of the conference and would like to make suggestions or proposals may contact Alain Azouvi at alain.azouvi@wanadoo.fr.

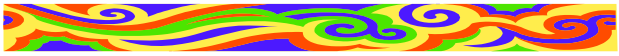
"Mini-courses" at Berlin

Elsewhere in this issue is the final list of the traditional "mini-courses" that will be given in conjunction with the Berlin Session. I would like to take this opportunity to thank our Scientific Secretary, Seppo Laaksonen, and those with whom he planned the program and the organizers of the Berlin Session, who are facilitating efforts to finalize the physical and practical aspects.

In Memoriam
Pierre Thionet
1915 - 2002

Pierre Thionet passed away on April 27, at the age of 87. Thionet studied at the École Normale, and became an associate professor in mathematics in 1942. He joined INSEE (Institut national de la statistique et des études économiques) in 1946 and was the architect of survey methods development in France. Drawn by this field's foundations in logic and mathematics, he wrote lecture notes and admirable contributions to the series "Études Théoriques de l'INSEE" during the 1950s. He defended a thesis on the loss of information in surveys in 1957, then left the world of administration for the academic one in 1959. He taught at Université de Poitiers and then at Université de Paris-X-Dauphine. Pierre Thionet was a survey visionary. His work contains—sometimes in embryonic form but often quite explicitly—all the major themes that have dominated survey research since the 1970s: the role of auxiliary information, weighting (today we would use the term "calibration"), the performance of constrained samples, robustness in relation to a working model, and probably other themes that I have not yet grasped. His thinking was ebullient and versatile. Every sentence in his work contains an idea (indeed, often more than one!). This did not, unfortunately, make his works easily accessible, especially for non-Francophones. We have lost one of France's greatest statisticians.

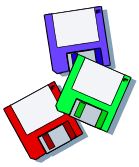
Prepared by Jean-Claude Deville



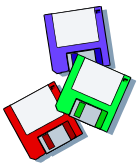
CHANGE OF ADDRESS

Members are encouraged to inform the IASS Secretariat of changes of address as soon as possible. Mailings of the forthcoming book of Leslie Kish's selected papers, proceedings of the IASS papers presented at the ISI sessions, and *The Survey Statistician* will be delayed and may be lost if the Secretariat does not have your correct address.

You may notify Ms. Claude Olivier of your change of address by completing and mailing the Change of Address form given at the end of this newsletter. Alternatively, you can provide the same information to Ms. Olivier by email to claude.olivier@insee.fr.



SOFTWARE REVIEW



g-CALIB 1.0: SPSS Based Software for Generalised Calibration

Camille Vanderhoeft
Statistics Belgium

1. Introduction

Vanderhoeft et al (2000) announced the existence of syntax modules g-Calib-S.sps and g-Design.sps developed for generalised calibration under SPSS®. Since then, we extended and improved the software in several respects, which finally has resulted in a tool that meets the following objectives: (1) being general with respect to methodology, thus incorporating basics and more specialised features that are useful for the practitioner; (2) being user-friendly, thus avoiding the practitioner having to be an expert in generalised calibration methodology; and (3) being well-structured, in order to allow flexible integration of the tool in the survey process.

The result is called g-CALIB, and its first release has recently become ready for dissemination.

The purpose of the present article is to promote g-CALIB 1.0, rather than to explain it in full detail. The latter would simply be beyond the scope of the present text; an extensive manual exists (Vanderhoeft, 2002) and we shall often refer to it. We also refer to our web site http://statbel.fgov.be/studies/cal_en.asp which provides links to quite a lot of information about calibration at Statistics Belgium, including examples that are mentioned in this article. In the next sections, we thus merely try to give a flavour of what kind of complexity a calibration problem might have, and how g-CALIB could be a useful means to solve such problems.

2. The Generalised Calibration Problem: Basics and Special Features

2.1. The Basic Problem

Given a sample *s* of size *n* from a population *U* of size *N*, a standard calibration problem can be defined as a constrained minimisation problem $\left\{min\ \mathrm{d}^T G(\mathrm{g});\ \mathbf{X}^T \mathrm{Dg} = \mathbf{t},\ \mathrm{g} \in \Omega_B\right\}$, where **X** is the $n \times m$ *calibration design matrix* containing information on *m auxiliary* or *calibration variables* for *n* sampled *elements*, **t** is a corresponding $m \times 1$ vector of *calibration totals*, **d** is a $n \times 1$ vector of *initial weights* (often the sampling weights, but a correction for non-response could already have been incorporated), $\mathrm{D} = diag(\mathbf{d})$, **g** is the (unknown) $n \times 1$ vector of *g-weights*, *G*(.) is a "*distance*" function, Ω_B occasionally defines explicit *boundary constraints* for the g-weights, and ^{T} indicates transposition of matrices and vectors. We can conveniently talk about a "calibration problem for elements given (auxiliary) information **X**,**d**,**t**". $\mathrm{X}^T\mathrm{Dg} = \mathrm{t}$ defines the system of *m calibration constraints*. Notice that $\mathbf{w} = \mathbf{Dg}$ is the $n \times 1$ vector of *calibrated weights*. For details (about the matrix notation, among other things), see Vanderhoeft (2001, 2002). See also Deville *et al* (1992, 1993), who introduced the generalised calibration problem and whose work has been the starting point for our work. An interesting result is that the g-weight for the *k*-th sampled element can be written $g_k = F\left(\mathbf{x}_k^T\boldsymbol{\lambda}\right) = g_k\left(\boldsymbol{\lambda}\right)$, where *F*(.), the *calibration function*, is the inverse of the derivative of *G*(.) and $\mathbf{x}_k^T$ is the row vector—called the *calibration vector*—of values of the *m* auxiliary variables for element *k*. Notice that different elements will have equal g-weights if their calibration vectors are equal, but if their initial weights are different they will have

different calibrated weights. g-CALIB 1.0 allows the user to choose the *calibration method*, i.e., the calibration function $F(.)$ or the distance function $G(.)$.

A *Newton-Raphson algorithm* is used to solve the constrained minimisation problem. If we define $\boldsymbol{\Phi}(\boldsymbol{\lambda}) = \sum_{k=1}^{n} d_k F\left(\mathbf{x}_k^T \boldsymbol{\lambda}\right)\mathbf{x}_k - \mathbf{t} = \mathbf{X}^T \mathbf{w}(\boldsymbol{\lambda}) - \mathbf{t}$ and $\boldsymbol{\Phi}'(\boldsymbol{\lambda}) = \mathbf{X}^T \mathbf{V}(\boldsymbol{\lambda})\mathbf{X}$, with $\mathbf{V}(\boldsymbol{\lambda}) = diag\left(d_k F'\left(\mathbf{x}_k^T \boldsymbol{\lambda}\right)\right)$, then successive updates for the Lagrange multipliers are calculated from $\boldsymbol{\lambda}^{(l)} = \boldsymbol{\lambda}^{(l-1)} - \left(\boldsymbol{\Phi}'\left(\boldsymbol{\lambda}^{(l-1)}\right)\right)^{-} \boldsymbol{\Phi}\left(\boldsymbol{\lambda}^{(l-1)}\right)$. Updates for the calibrated weights are $\mathbf{w}\left(\boldsymbol{\lambda}^{(l)}\right) = \mathbf{D}\mathbf{g}\left(\boldsymbol{\lambda}^{(l)}\right) = \mathbf{D}F\left(\mathbf{X}\boldsymbol{\lambda}^{(l)}\right) = \mathbf{D}\mathbf{g}^{(l)}$. Starting values are always $\boldsymbol{\lambda}^{(0)} = \mathbf{0}$. Iteration stops when $\max_{1 \le k \le n} \left| g_k^{*(l)} - g_k^{*(l-1)} \right| \le \varepsilon$, for a given tolerance ε, or when a pre-specified maximum number of iterations is reached. For the linear method, iteration already stops after the first step. The superscript * results from the adaptation of the classical Newton-Raphson algorithm to solve calibration problems with additional explicit boundary constraints for the truncated linear method. We refer to Vanderhoeft (2002, section 2.4) for details.

g-CALIB 1.0 uses the same technique as is used in Calmar (see Deville *et al*, 1993 and Sautory, 1993) for solving the calibration problem, including for the case of the truncated linear method. However, g-CALIB works with *generalised inverse matrices*, which means that redundant constraints need not be removed; Calmar can only handle ordinary inverse matrices. Bascula 4.0 (Nieuwenbroek & Boonstra, 2002a and b) does not use a general algorithm, but distinguishes between the following cases: poststratification, ratio estimation, linear and multiplicative weighting. Different algorithms have been implemented in Bascula 4.0 for these distinct situations. This makes Bascula 4.0 less general than both g-CALIB 1.0 and Calmar. Moreover, we have implemented several special features in g-CALIB 1.0, resulting in a package that seems to demand less advanced technical skills from the practitioner than if s/he would use Calmar.

2.2. Calibration Strata

Suppose a sample of individuals has to be calibrated to a (target) population, taking the (joint and/or marginal) distribution of the population relative to three categorical variables into account: PROVince, AGE-classes and SEX. As in multiple regression modeling, the structure of the design matrix defines the *linear structure* of the calibration model, and can be efficiently described by symbolic expressions such as PROV * AGE * SEX, PROV + AGE + SEX, PROV*AGE + SEX, AGE*SEX, PROV*(AGE + SEX), etc.

Consider the linear structure PROV*(AGE + SEX). This symbolic expression implies a design matrix with a particular block-diagonal structure (provided that the sampled individuals are ordered with respect to the variable PROV). It can then be shown that solving the calibration problem using the whole sample is equivalent to separately solving a reduced problem for each category of PROV. The user of g-CALIB can exploit this feature, using PROV as a so-called *calibration stratification variable*. The result is that g-CALIB performs the calibration separately within each calibration stratum (a category of PROV), taking them one after another, and assuming the linear structure AGE + SEX in each of them.

2.3. Nested Levels of Auxiliary Information

A standard calibration problem will involve a sample of, say, *elements*—often the statistical units of primary interest—and values for auxiliary variables at the level of the elements in that sample. However, auxiliary information might also be available for *clusters* of sampled elements (and hence for each element in the cluster!). If elements are individuals, clusters often will be households or families. Typical cluster variables are: household size, number of children in the household, professional status of the reference person, total expenditure of the household, etc. g-CALIB 1.0 allows the user to work with 2 nested levels of auxiliary information (see Vanderhoeft, 2002, section 2.2.6 for 3 or more levels).

Level of calibration	The calibration problem transformed to a "standard" optimisation problem	g-Weights
Level 1 : Elements, no clustering	$\left\{ min\ \mathrm{d}^T G(\mathrm{g});\ \mathbf{X}^T \mathrm{Dg} = \mathbf{t},\ \mathrm{g} \in \Omega_B \right\}$	$g_k = F\left(\mathrm{x}_k^T \lambda\right)$
Level 2 : Elements, with clustering	$\left\{ min\ \mathrm{d}^+ G(\tilde{\mathrm{g}});\ \tilde{\mathrm{H}}^T \mathrm{D}^+ \tilde{\mathrm{g}} = \mathbf{t},\ \tilde{\mathrm{g}} \in \tilde{\Omega}_B \right\}$	$\tilde{g}_l = F\left(\bar{\mathrm{x}}_{(l)}^T \lambda'\right) = g_k$
Level 3 : Clusters	$\left\{ min\ \tilde{\mathrm{d}}^T G(\tilde{\mathrm{g}});\ Z^T \tilde{\mathrm{D}} \tilde{\mathrm{g}} = \mathbf{s},\ \tilde{\mathrm{g}} \in \tilde{\Omega}_B \right\}$	$\tilde{g}_l = F\left(\mathrm{z}_l^T \gamma\right)$
Level 4 : Elements + Clusters	$\left\{ min\ \tilde{\mathrm{d}}^T G(\tilde{\mathrm{g}});\ \begin{pmatrix} Z^T \\ \hat{\mathrm{X}}^T \end{pmatrix} \tilde{\mathrm{D}} \tilde{\mathrm{g}} = \begin{pmatrix} \mathrm{s} \\ \mathrm{t} \end{pmatrix},\ \tilde{\mathrm{g}} \in \tilde{\Omega}_B \right\}$	$\tilde{g}_l = F\left(\mathrm{z}_l^T \gamma^* + n_l \bar{\mathrm{x}}_{(l)}^T \lambda^*\right)$ $= F\left(\mathrm{z}_l^T \gamma^* + \hat{\mathrm{x}}_{(l)}^T \lambda^*\right) = g_k$

Given information at both levels, i.e., **X**,**d**,**t** at element level and, similarly, $\mathbf{Z},\tilde{\mathbf{d}},\mathbf{s}$ at cluster level, the user can choose among four *levels of calibration*, depending on the desired result and on the use of calibrated weights for estimation purposes. The above table gives an overview of the optimisation problems, and the corresponding expressions for the g-weights, that result for each level of calibration after some transformations. Details are found in Vanderhoeft (2001, 2002); we briefly comment on each calibration level.

Level 1: Element level calibration, without clustering

This is a standard calibration problem, given element information **X**,**d**,**t**. It has been discussed in section 2.1. Each element k is finally assigned a g-weight g_k and a calibrated weight $w_k = d_k\, g_k$.

After calibration of the sample of elements, totals for element level variables can be estimated easily using the resulting element level calibrated weights. But what should be done if estimates of cluster characteristics are required? It is common practice then (e.g., so far for the Labour Force Survey in Belgium) to assign a weight to each cluster in one of the following ways: 1) assign the calibrated weight of a representative element (e.g., the reference person if clusters are households) in each cluster to the cluster, or 2) assign the average of the calibrated weights of all elements in the cluster to the cluster. These *ad hoc* procedures have no rigorous mathematical justification, and often seem to fail in practice.

Level 2: Element level calibration, with clustering

The ad hoc methods can be avoided by incorporating the additional constraints that elements in the same cluster should have the same g-weights. It can be shown that the calibration problem still can be transformed into the basic form (as in section 2.1) as shown in the above table, where the superscript $^+$ indicates summation within clusters and the elements of $\tilde{\mathbf{H}}$ are weighted averages of element calibration variables within clusters. Reduction to cluster level—a procedure called *clustering*—is indicated by the $\tilde{}$ sign. The resulting g-weights are appropriate at both cluster and element level. Calibrated weights follow easily.

So far, we have used element (individual) level auxiliary information only for benchmarking. This generally affects the quality of estimates of cluster (household) characteristics. Level 3 and Level 4 calibration might help to overcome that problem.

Level 3: Cluster level calibration

Benchmarking on cluster level population figures **s**, such as the number of households, is possible if cluster level information $\mathbf{Z},\tilde{\mathbf{d}},\mathbf{s}$ is available. This is a standard problem (for clusters instead of elements).

Level 3 calibration is expected to provide higher numerical consistency at cluster level, but benchmarking at element level is now completely ignored. Estimates of element level characteristics might thus be less consistent with known information. Notice that element weights can be derived from cluster weights in several *ad hoc* ways, similar to the problem of getting cluster weights after level 1 calibration.

Level 4: Simultaneous element-cluster level calibration

Benchmarking simultaneously at element level information $\mathbf{X},\mathbf{d},\mathbf{t}$ and cluster level information $\mathbf{Z},\tilde{\mathbf{d}},\mathbf{s}$ is possible and preferable if both element and cluster level characteristics have to be estimated. The idea is to transform all information to cluster level and calibrating the cluster sample: see the above table. Notice that the matrix $\widehat{\mathbf{X}}$ contains sums of element level calibration variables within clusters. The assumption $d_k = \tilde{d}_l$ if $k \in l$ is made. The resulting g-weights are valid for elements and for clusters.

2.4. Other Technical Features: Collapsing and a Scale Parameter

Collapsing (Vanderhoeft, 2002, section 2.3.2) is a technical feature, implemented for the purpose of speeding up calculations, i.e., to improve performance of the algorithm. The technique exploits the fact that g-weights are equal for elements or clusters that have identically the same calibration vectors (after transformation, depending on the level of calibration). Elements or clusters with the same calibration vector are said to belong to the same *collapsing cell*, their data are aggregated and the basic problem of calibrating a set of elements or clusters is transformed into a reduced problem of calibrating a set of different collapsing cells. The user of g-CALIB 1.0 only has to be aware of this feature for correct interpretation of part of the informative output. Numerical output is not affected by collapsing.

g-CALIB 1.0 further allows the user to calculate a global adjustment (per calibration stratum), occasionally an overall correction for non-response represented by the *scale parameter* ϕ of the calibration problem. The user can set this feature on or off. If the option is set off, the user can assign a fixed value to ϕ (the same for all calibration strata). The g-weights produced by g-CALIB 1.0 are always relative to the *scaled weights* $\mathrm{d}_\phi = \phi \cdot \mathrm{d}$. The user should be aware of the fact that the results depend on the value of ϕ, except if the exponential method is used.

3. The Software

3.1. Some Important Windows: an Example

The core modules of g-CALIB are parameterised SPSS® syntax files, which are run through an SPSS® Production Facility job. All this is further managed through g-CALIB's user-friendly Windows-like interface. The latter is developed under Visual Basic. The windows displayed by this interface allow the user to define, modify, load, and execute his/her proper calibration jobs.

Figure 1. Main window for job Linear-E.gc1

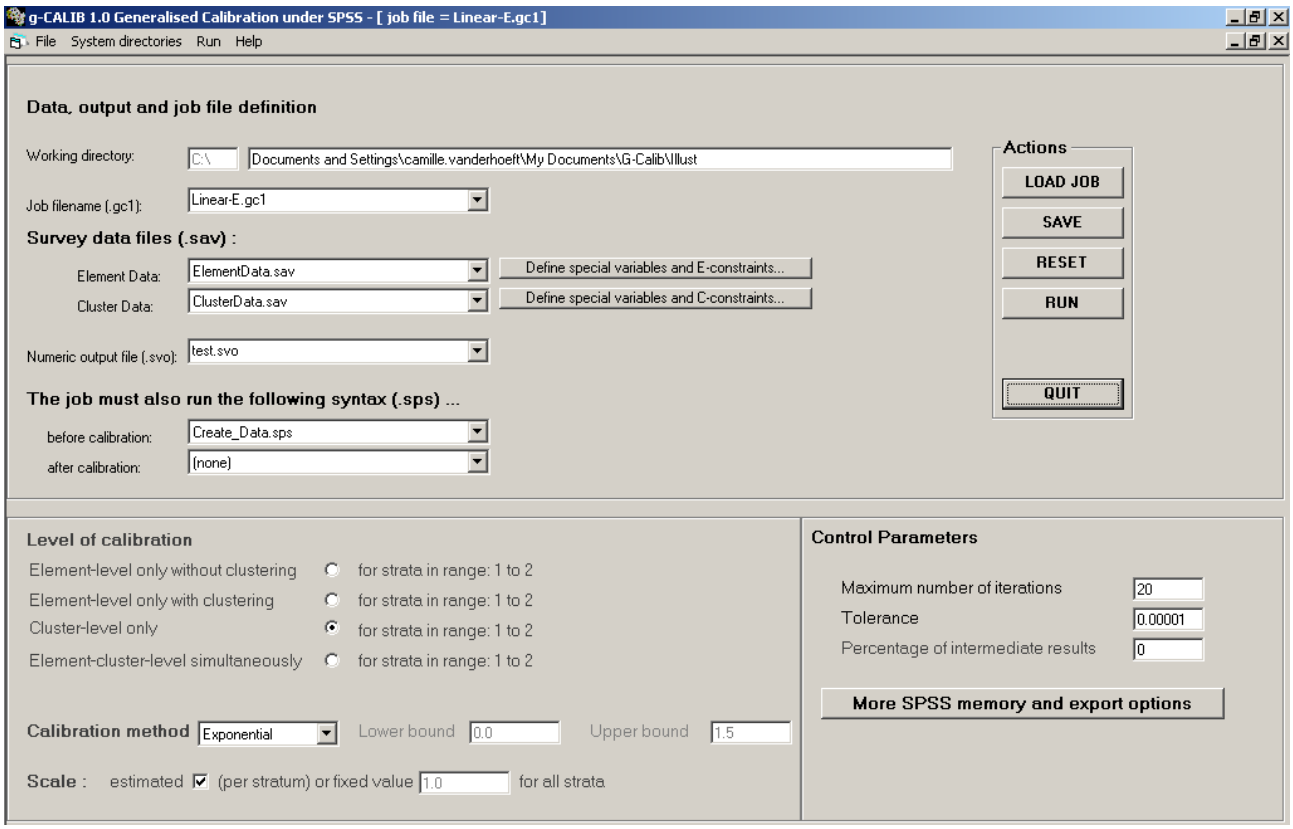
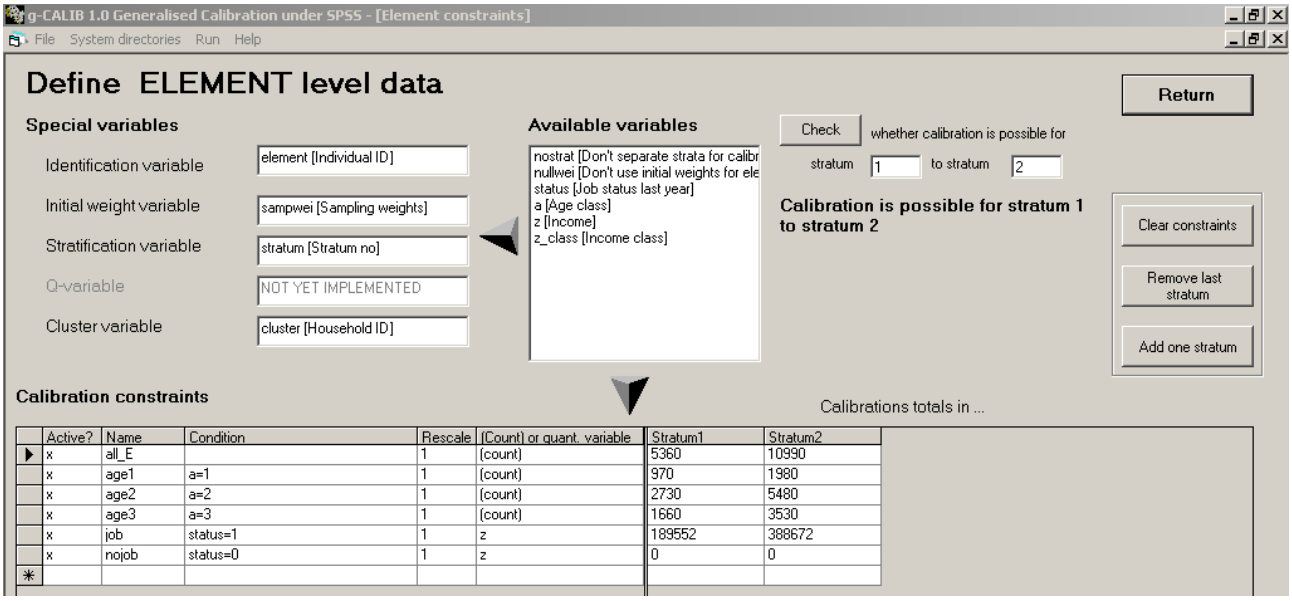


Figure 1 shows the main window after loading a calibration job called *Linear-E* (or Linear-E.gc1). This job is one of the examples—based on hypothetical data—that are delivered with the software and discussed in the manual (Vanderhoeft, 2002). The .gc1 file, and other files related to this job (including a short description of the problem), can be downloaded from the section *Applications* (no. 00.02) on our web page http://statbel.fgov.be/studies/cal_en.asp. The window in Figure 1 consists of several panels: panel 1 wherein the user defines location and names of input and output files; panel 2 for defining the level of calibration, the calibration method, the scale parameter; panel 3 for setting control parameters (with a gateway to more advanced settings); and panel 4 with some action buttons. The user is referred to other documents for more details about how to use this window to define or modify a calibration job.

Clicking the *Define special variables and E-constraints…* or *Define special variables and E-constraints…* button in panel 1 opens a new window, shown in Figure 2 (for elements) for job Linear-E. Special variables and calibration constraints are easily defined by dragging and dropping, typing, copying and pasting, etc. Some special features allow the user to activate/inactivate calibration constraints (the first column in the grid), to rescale constraints for better performance in case of numerically hard problems, to add/remove calibration strata, to select a range of strata for which calibration should be carried out, etc. The software checks whether active constraints are well-defined, gives an informative message and if necessary proposes another range of calibration strata.

Figure 2. E-constraints window for job Linear-E.gc1



	Active?	Name	Condition	Rescale	(Count) or quant. variable	Stratum1	Stratum2
▶	x	all_E		1	(count)	5360	10990
	x	age1	a=1	1	(count)	970	1980
	x	age2	a=2	1	(count)	2730	5480
	x	age3	a=3	1	(count)	1660	3530
	x	job	status=1	1	z	189552	388672
	x	nojob	status=0	1	z	0	0
*							

Given that the classes “a=1”, “a=2” and “a=3” are exhaustive, and that the last constraint is of type 0=0, since z=0 for all elements in class status=0 and since the corresponding calibration totals too are 0, there are 2 redundant constraints, which could be but need not be inactivated (or removed). This is a consequence of using generalised inverse matrices (section 2.1): the design matrix need not have full rank; it causes the software to be very flexible and user-friendly with respect to defining constraints. These features can also be used to easily switch from one model to another, thus changing the set of active constraints.

3.2. Input and output files

g-CALIB’s survey data input files can be very elementary SPSS® data files. Only a few rules must be satisfied: 1) elements or clusters need to be identified (within calibration strata) by a numeric key; 2) an initial weight variable must be present, even if no initial weighting is considered (the variable is then constant with value 1); 3) a calibration stratification variable (numbering strata as 1, 2, …) must be present, even if no strata are considered (the variable is then constant with value 1); 4) the stratification variables in element and cluster data files must have the same names if calibration level 4 is chosen; and 5) the element data file also has to contain the cluster identification variable, with the same name, for linkage of element data and cluster data files if calibration level 4 is chosen.

Other variables can be anything; some of them may be used to define calibration constraints, while others will not be used at all for calibration. Notice that the mathematical formulation in the previous section implies that the columns of a design matrix are often indicator variables, taking values 0 and 1, to identify elements or clusters in a particular category of a qualitative variable (such as PROV, SEX, …). Such indicator variables can, but need not be calculated by the user or be present in the input data files. They will be internally constructed by g-CALIB, based on the user’s definition of the calibration constraints, when a calibration job starts running.

With respect to calibration totals, it is generally supposed that they are available when the user starts defining a calibration job. They are usually entered one after another through g-CALIB’s interface (see section 3.1), and stored by g-CALIB in the *job definition file* (a file with extension .gc1, which is merely a text file).

Output from g-CALIB 1.0 is twofold: extensive *informative output* in a SPSS® viewer document (a file with extension .spo) and *numerical output* in a SPSS® data file (a file with extension .svo, to distinguish them from other SPSS data files that usually have extension .sav).

The .spo file contains lots of information on the sample (e.g., sample totals of calibration variables), on calibration totals, on the calibration model chosen, on the iterative procedure (number of iterations, rank of the design matrix, number of negative weights, etc.), characteristics of the distribution (by calibration stratum) of

scaled, g- and calibrated weights, box-plots (by calibration stratum) of weight variables, a table for comparison of various totals (fixed, initial, scaled and calibrated), etc. The reader is referred to the manual (Vanderhoeft, 2002, chapter 7) for detailed discussion of a .spo file. The information in this file is particularly useful to check whether the software has run properly and whether the numerical output is trustworthy for other purposes (e.g., estimation).

Numerical output is stored in a .svo file, which is an ordinary SPSS® data file. This file contains: the identification variable for elements or clusters, the calibration stratification variable, the initial weight variable, the full design matrix for the current model, a variable called SCALE containing fixed or estimated values for the scale parameter ϕ, a variable called SCAWEI containing the scaled initial weights, a variable called CALWEI containing the final calibrated weights, and a variable called G_WEIG containing the final g-weights. Using the identification variable (with the user's original name), the .svo file can easily be linked with the user's original input file(s). This can be the starting point for estimation of totals (and other parameters) of survey variables that are present in the original input file.

g-CALIB 1.0, while executing a calibration job, creates several other SPSS® data files, which are not erased at the end of the job. These intermediate files could thus be used for special purposes, though it will require some more effort from the user. Being familiar with SPSS® programming features is an asset in these cases. Details about these files are found in the manual (Vanderhoeft, 2002, section 5).

3.3. Advanced Features: Flexibility in Defining Constraints and Extended Calibration Jobs

We discovered g-CALIB to be useful not only for calibration of micro data (from sample surveys), but for adjustment of cross-tabulated (or macro) data as well. When dealing with input-output matrices, the RAS method, for instance, adjusts the cells in a table of transactions to fixed row and column totals. This is a raking problem, which can be solved with g-CALIB, using the exponential method. The trick is to construct a data file that contains the data in the input-output table in a format that can be handled by g-CALIB: the cell entries (transactions) become the initial weights; a cell numbering provides an identification variable; a calibration stratification variable can be a numbering of different input-output tables, such that several tables can be adjusted independently in a single run of g-CALIB; variables identifying rows and columns will allow the user to construct row and column constraints.

Active?	Name	Condition	Rescale	(Count) or quant. variable	Stratum1
x	Row1	ROW=1	1	(count)	1000
x	Row1bis		1	ROW=1	1000
x	R1_C1		1	(ROW=1)–(COL=1)	0

The above table shows a grid (as in Figure 2) defining three constraints, called Row1, Row1bis and R1_C1. Given that the variable ROW indicates to which row a cell belongs, Row1 and Row1bis are equivalently fixing the total for row 1 in the table. Constraint R1_C1 is contrasting the first row and the first column. Since the numeric result of the expression in the fifth column is 1 if ROW=1 and COL$\neq$1, –1 if ROW$\neq$1 and COL=1, and 0 otherwise, and since there is no condition, the underlying calibration variable (a column in the design matrix) finally takes the values 1, –1 and 0, which is called a *contrast variable*. Notice the value 0 for the fixed total for this particular constraint. This constraint causes the totals for row 1 and column 1 in the table to be equal after adjustment. More about this is found on our web site in the *Applications* section (examples 02.**).

Another advanced feature consists of extending a calibration job by defining a "before calibration" syntax file, which is executed before any other core module, and an "after calibration" syntax, which is executed after any other core module. This allows the user to flexibly integrate calibration into the entire survey process. The main window in Figure 1 shows that the job Linear-E is such an extended job: it includes a syntax called Create_Data.sps; see the manual (Vanderhoeft; 2002) for details. Pull-down lists allow selecting the files from the list of .sps files in the current working directory.

We have used these possibilities of extending a job extensively in our work on real surveys. A "before calibration" syntax has, for instance, been developed for imputation of an incomplete data set from the Labour

Force Survey. “Before calibration” syntaxes will often be used to prepare the final input file(s) from which calibration actually starts. We have also integrated “after calibration” syntaxes for estimation of totals of study variables, for estimation of variances using the random group method, and for tabulating adjusted (input-output) tables. The manual (Vanderhoeft, 2002, chapter 8) suggests some other uses of “before” and “after calibration” syntaxes.

4. How to Obtain g-CALIB 1.0

On our web site (http://statbel.fgov.be/studies/cal_en.asp for the English version) the interested statistician can find the link “How to Obtain g-CALIB 1.0?” through which a document presenting our dissemination policy and a form for ordering the software can be downloaded. The document should be completed and sent to the author. Support during installation can be provided as needed.

For more specific information please contact:

Mr. Camille Vanderhoeft
Statistics Belgium - Department of Methodology and Coordination
Leuvenseweg, 44, B-1000 Brussels,
Belgium
Fax: +32 3 548.66.90
email: camille.vanderhoeft@statbel.mineco.fgov.be

References

Deville, J.-C. and Särndal, C.-E. (1992) Calibration Estimators in Survey Sampling, *Journal of the American Statistical Association*, Vol. 87, No. 418, pp.376-382.

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

Nieuwenbroek, N. and Boonstra, H.J. (2002a) *Bascula* 4.0 *Reference Manual*. Statistics Netherlands.

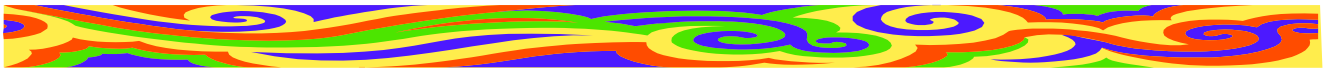
Nieuwenbroek, N. and Boonstra, H.J. (2002b) Bascula 4.0 for Weighting Sample Survey Data with Estimation of Variances. *The Survey Statistician*, IASS Newsletter, No. 46, pp 6-11.

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

Vanderhoeft, C., Museux J.-M. and Waeytens, E. (2000) g-DESIGN and g-CALIB-S : SPSS® modules for Generalised Calibration. *The Survey Statistician*, IASS Newsletter, No. 43, pp. 6-12.

Vanderhoeft, C. (2001) Generalised Calibration at Statistics Belgium. SPSS® Module g-CALIB-S and Current Practices. *Statistics Belgium Working Paper* no.3.

Vanderhoeft, C. (2002) *g-CALIB Release* 1.0 - *Generalised Calibration under SPSS®*. Statistics Belgium.



.

BRAZIL

from Pedro Luis do Nascimento Silva

In 2003, the **"Escola Nacional de Ciências Estatísticas"** (ENCE) from the Brazilian Central Statistical Office (IBGE) will complete 50 years of service. As part of commemorations, a number of events will take place, starting from March onwards. ENCE was the first institution to provide undergraduate degree (BSc.) in Statistics in Brazil, since 1953. Following its tradition of innovation in statistical education in Brazil, it started offering from 1998 a masters program in Population Studies and Social Surveys, with three concentration areas: Demography, Social Statistics; and Population, Society and Territory. Completion of the program should take two years for full time students, and over 15 dissertations have already been successfully completed. Further information available from: http://www.ence.ibge.gov.br/.

Curso de Desenvolvimento de Habilidades em Pesquisa (CDHP) is a training program that aims to enable staff at IBGE to plan and carry out (household) sample surveys. This course was developed taking as the model the Survey Skills Development Course of Statistics Canada. It presents key concepts, methods and steps required to successfully develop and implement a household sample survey. This is accomplished by means of lectures, guided teamwork and do-it-yourself activities that include developing and carrying out a real-life survey experience. Each course "crew" comprises 25 trainees, plus 10 to 15 tutors (some part-time), alongside a course coordinator and administrative assistant. Each course has a genuine survey client, carries out a survey to satisfy this client's demands, and delivers the outcome in the form of traditional survey products: a survey report, survey documentation and data files, plus a seminar-like presentation of main results, which is attended by IBGE's directing board, staff and the client. Each course runs for six weeks full time. Data collection activities are performed solely by the crew of trainees each time. Nine courses were successfully developed since 1997, with surveys covering various topics of interest. The course has proved very valuable as a staff development, integration and motivation strategy, plus provided a number of relevant survey outcomes. Flexibility, focus on outcome and client satisfaction, teamwork and transmission of corporate values, goals and feeling of membership are some of the high points, which make this a very successful course amongst both staff and managers. Future plans being considered include opening up vacancies for people who are not IBGE employees, as well as extending capabilities for running three courses a year. Further information is available from http://www.ence.ibge.gov.br/.

Inter-American Statistical Association (IASI) is a professional organization that aims to promote the statistical development in the American region. As part of its program of activities, IASI has regularly carried out a series of meetings called "*Seminars of Applied Statistics*". The latest of these seminars, held in July 2001 in Panama, dealt with "*Statistical Methods for Quality and Productivity*". The next seminar in this series is to take place in Rio de Janeiro, Brazil, from 7 to 10 July 2003. The themes chosen for the *IX Seminar of Applied Statistics* are "*Statistics in Education and Education in Statistics*". This seminar should provide an excellent opportunity for exchange of ideas, dissemination of recent work and developments that took place in Brazil and the Americas over the last few years, together with discussion of perspectives for advancement of both areas in the future. The goals are to attract wide participation from researchers, university teachers, and professionals who have an interest in the themes of the seminar from Brazil and other countries, as well as to encourage participation from students (graduate and undergraduate) and high-school teachers. The Seminar will feature short courses, invited conferences, round tables, oral presentations and poster session, amongst other activities. Deadline for paper submission and discount registration fees is March 31, 2003. Further information about the seminar is available from: http://www.indec.mecon.gov.ar/iasi.htm.

Analysis of complex survey data using free software has received a boost in Brazil. Prof. Djalma Pessoa (djalma@ibge.gov.br) developed a suite of functions using the statistical software R (http://www.r-project.org/) that can be used to perform survey estimation and analysis tasks, such as estimating means, totals, proportions, ratios, linear and logistic regression models, as well as corresponding sampling errors and test statistics, taking into account the complex survey design. The software is based on Taylor linearisation methods, runs under the Windows implementation of R, has some limited documentation in the English language, and is available free of charge from the author as an R package that can be easily installed and utilized by the user.

CANADA

from John Kovar

The data capture of the 2001 **Canadian Census of Agriculture** was conducted between July and November 2001, using relatively new technology called **Intelligent Character Recognition** (ICR). The ICR system offered many benefits to the data capture operation, in terms of resource savings and productivity gains. At the same time, accuracy was an extremely important consideration since the potential for unacceptable levels of error existed at various stages of the process. Therefore, quality assurance (QA) and quality control (QC) procedures were built into this operation to ensure a high degree of accuracy in the data capture process. Quality activities were focused in three main stages of processing, namely: document preparation, scanning & recognition, and the actual data capture of the questionnaires.

It was estimated that about 77 percent of the fields were captured through the automated data capture process with the balance captured by Key-from-Image processing. The ongoing application of QA/QC, along with system adjustments and feedback, resulted in the substitution error rate dropping from 4.1 percent at the beginning of the process, to 0.8 percent during the last week. The overall substitution error rate for the entire process was measured at 2.0 percent on a field basis for all field types. It is clear from these results that the QA/QC procedures were extremely valuable and had a positive impact on the entire operation. The QA procedures that were applied in the Document Preparation process were effective in preventing many poor documents from reaching the scanning stations and those that did were then labeled for special treatment and subsequent 100 percent verification. The QC procedures were then able to optimize the machine set-up by applying the Scanning Calibration Check prior to production. Furthermore during production, QC samples were able to identify problems with the automatic data capture and key from image processes, so that they could be improved as required. In all cases, early warning signals were obtained from objective measurements at each stage, and corrective and preventive actions were implemented as needed.

For further information regarding the quality assurance and quality control aspects of this project, please contact Walter Mudryk (walter.mudryk@statcan.ca or at 613-951-4781).

Many minimum change imputation systems are based on the approach proposed by Fellegi and Holt. For example, CANEDIT and GEIS at Statistics Canada, and DISCRETE and SPEER at the United States Bureau of the Census all use, or had as their starting point, the Fellegi/Holt imputation methodology. In the **1996 Canadian Census of Population**, a somewhat different approach was used successfully to impute for nonresponse and resolve inconsistent responses for the demographic variables of all persons in a household simultaneously. The method used is called the **Nearest-neighbour Imputation Methodology (NIM).** This implementation of the NIM allowed, for the first time, the simultaneous hot deck imputation of qualitative and quantitative variables for large Edit and Imputation (E&I) problems.

The main difference between the NIM and the Fellegi/Holt imputation methodology is that the NIM first finds donors and then determines the minimum number of variables to impute based on these donors. The Fellegi/Holt methodology determines the minimum number of variables to impute first, and then finds donors. Reversing the order of these operations confers significant computational advantages to implementations of the NIM while still meeting the well-accepted Fellegi/Holt objectives of minimum change and preserving sub-population distributions. The NIM, however, in its present form, can only be used to carry out imputation using donors while the Fellegi/Holt can be used with any imputation methodology.

For the **2001 Census**, a more generic implementation of the NIM has been developed. It is called the **CANadian Census Edit and Imputation System (CANCEIS).** Besides the demographic variables, it was used in the 2001 Canadian Census to perform E&I for the labour, mobility, place of work, and mode of transportation variables. This corresponds to about half of all variables on the 2001 Census questionnaire. All processing was carried out on PCs. For the 2006 Canadian Census, CANCEIS will be used to process all census variables. In addition, CANCEIS has been used by the Canadian Census of Agriculture Coverage Evaluation Survey and the Canadian Survey of Household Spending.

CANCEIS (or an earlier version of the software) has been or will be used to process some of the variables in the 2001 Ukrainian Census, the 2000 Brazilian Census and the 2001 Swiss Census. In addition, the 2001 Italian Census, having studied CANCEIS, will use a similar approach in their imputation methodology.

For further information on CANCEIS, please contact Mike Bankier (bankier@statcan.ca or at 613-951-6938).

HUNGARY

from Laszlo Mihalyffy

Redesigning the Samples of Household Surveys on the Basis of the 2001 Population and Housing Census in Hungary. Preparations for and the carrying out of the 2001 Population and Housing Census in Hungary were described in earlier country reports. In this note, we report on the redesign of the samples of the Household Budget Survey (HBS) and the Labour Force Survey (LFS) as well as designing two new surveys on the basis of that census.

Up to now, the sample design of the household surveys conducted by the Hungarian Central Statistical Office (HCSO) has been based on the following principles: the primary sampling units (PSUs) are census enumeration districts (EDs) and localities in the self-representing (SR) and in the non-self-representing (NSR) parts of the sample, respectively. The ultimate sampling units (USUs) are dwellings in each case, while in the NSR part of the sample, EDs are the intermediate sampling units. This scheme will partly be changed, since an address register has been created from the list of addresses of the 2001 Census, which supports, for example, the selection of dwellings with systematic sampling in the first stage in the SR part of the sample. The most common stratifying variables are the size of the localities and the administrative units (e.g., counties). Some basic facts on the current and future HBS and LFS samples are set forth in tabular form.

	Samples			
	HBS (Annual)		LFS (Quarterly)	
	Current	Future	Current	Future
No of Localities	262	273	754	657
Of Which: Those in the SR Part	77	90	126	130
No of Census Enumeration Districts	1,977	1,927	12,829	–
Of Which: Those in SR Localities	1,072	1,224	6,295	–
No of Addresses	11,862	11,562	38,487	37,713
Of Which: Those in SR Localities	6,432	7,344	18,885	18,033

In contrast to the future HBS where the EDs in the sample play an important role in organising data collection, EDs are not used as sampling units in the future LFS. Using preliminary census information, the allocation of the future LFS sample was determined to comply with the Eurostat recommendation requiring that the coefficient of variation (CV) of the estimated level of unemployment for counties should not exceed a given limit. In addition, the designed sample weights would also be kept within reasonable limits. In the future HBS sample, households will be stratified by the activity of the head. Using higher and lower sampling rates for households with head of third level education and those with head aged 60 and over, respectively, is going to reduce the impact of nonresponse on the structure of the sample.

The Survey on the National Tourism Demand (NTD) will be launched in the second half of 2003. Utilizing the experience of a similar Spanish survey, the quarterly NTD will use a sample similar to the future LFS sample; the sample size will amount to 12,000-13,000 households. The annual Survey on the Possession and the Use of Informatics and Telecommunication (ITC) Devices and Services by the Households will also be launched in 2003. While the sampling frame for that survey, too, will be the address register, both the survey and its sample will be rather specific, since the ITC devices and services are distributed rather unevenly among the households of different

regions of the country. Some strategic questions of sampling concerning this survey have still to be solved.

For further details please contact the authors: Ödön Éltető, odon.elteto@office.ksh.hu, László Mihályffy, laszlo.mihalyffy@office.ksh.hu, Ákos Probáld, akos.probald@office.ksh.hu.

ISRAEL

from Natalie Shlomo and Luisa Burck

The Israel Central Bureau of Statistics (CBS) is in the process of developing a comprehensive **Statistical Business Register** to be established by the year 2004. The Business Register will consolidate the various administrative business files currently in use at the CBS, comprising mainly of the VAT (Value Added Tax) and National Security Agency files. The Business Register will serve as the unified source of business data and will be defined, maintained, and updated under the responsibility of the CBS. The Business Register will provide ongoing demographic data and business indicators for the business entities, including new businesses, mergers, splits, and inactive or dormant businesses. Besides the cross-sectional data, the Register will also provide a source for longitudinal business data. To build the Business Register, administrative files will be linked using different matching techniques. Both Economic Units and Statistical Units will be defined and classified, and a unique ID number will be assigned to each unit. The main purpose of the Business Register from a methodological standpoint is to provide sampling frames for the various monthly and yearly business surveys currently being carried out at the CBS. The Register will be used to plan and co-ordinate the samples. The surveys are longitudinal where the panel usually lasts for about 5-7 years. Samples are drawn with probability proportional to an estimated size within strata defined by industries. All historical data on the samples will be maintained on the Register so that each sample can be updated with minimum changes to the original sample. For more information, contact Tzahi Makovky, Statistical Methods Sector, Central Bureau of Statistics, Israel, Tel: (972)-2-659-2724 tzahim@cbs.gov.il.

The Central Bureau of Statistics in Israel is in the process of implementing **X12-ARIMA** as the standard method of seasonal adjustment. X12-ARIMA (Findley et al., 1998) retains the general approach of X-11-ARIMA/88 in seasonal adjustment calculations, with many enhancements that include major improvements to the pre-adjustment and post-adjustment phases. The estimation of the moving Jewish festivals and the trading day effects has always posed a difficult problem in Israel. Econometric modeling and seasonal adjustment analyses using monthly data are difficult unless the effects of these moving holidays can be estimated with some precision. For the estimation of these effects, the Statistical Analysis Sector developed a special method. Studies showing how this method can be implemented using X12 have been carried out. Furthermore, we expanded the method to include detection and estimation of trend breaks as well as the automatic outlier detection processes using regArima models. For more information, contact Luisa Burck, Statistical Analysis Sector, Central Bureau of Statistics, Israel, Tel: (972)-2-659-2742 louiza@cbs.gov.il.

ITALY

Claudio Quintano

Process innovations in the demographic and social statistics by means of administrative sources. The need to drastically reduce information release times and to strengthen and consolidate the data quality improvement process on the one hand, and the strong technological innovations and the dissemination of network technologies including the Internet (also within the Public Administration) on the other hand, have brought about a general reorganisation of the whole production process. At the National Statistical Institute of Italy (ISTAT) this affected the processes by which micro and macro data are collected, disseminated and analysed.

DATA CAPTURING

An important innovation is the automatic collection of data from Population Registers and Civil Status sources, using the *ISTATEL* service with the development of a software named "*ISI-ISTATEL*" that all the Municipalities can use to collect, check and send data on vital statistics to a service centre, or by filling in the electronic survey questionnaires that exactly reproduce the paper forms or uploading data from standard format files. The data

collected in this way are checked and, if they fulfil preset criteria for consistency and correctness, are entered into a single database. The surveys used within the ISTATEL service are mainly based on individual data, such as those on births, marriages, deaths and changes of residence. ISTAT, however, carries out three annual surveys on resident population at the Population Registers' Office in each Municipality. Data collected refer both to the whole population (flows and stock data with the distribution by sex, age and marital status), and the foreign component, with the distribution by sex and citizenship. For these three surveys an electronic version of the forms (*MODEM* project) has been created on a web platform using a Computer Assisted Web Interviewing (CAWI) technique; the Municipalities have therefore been given the possibility of choosing whether to use the traditional paper form, or the new method of filling in and sending data via internet using the ISTAT web data collection site. Generally, the advantages provided by using methodologies of automatic data capturing are a reduction and progressive elimination of the amount of paper questionnaires and filling in of the latter by hand; the possibility to enter consistency and correctness checks directly at source, with operators' possibility to remove errors; the use of online guides for filling in forms; the control of the questionnaire's route, with consequent reduction of the interview (the person interviewed can be asked only the parts of the questionnaire that must be answered in reference to other filter questions); the absence of sending times; the reduction of costs linked to printing, sending, recording and editing of data; the possibility of creating user-friendly applications with pleasant graphic interfaces which do not require any particular computer knowledge by the user; and the independence from hardware platforms since a web questionnaire can be accessed by any personal computer that has an Internet connection and a browser.

DATA DISSEMINATION

In order to satisfy the growing demand for statistical information on demographic issues in real time, and to provide data at the greatest possible level of detail, the *DEMO-Population and Demographic Statistics* web warehouse, currently operating at the address http://demo.istat.it, has been designed and implemented. The published data are the result of the three annual surveys on resident population carried out by ISTAT at the register offices of the single municipalities. Recently the Projections of Population by age, sex and region of residence for the years 2001-2051 and the life tables of the Italian population by province and region of residence have been added. DEMO is a system for data query and dynamic creation of tables by the user. Indeed, it makes it possible to choose, for each single survey, a specific territorial level (the whole of Italy or a district, region, province, or even a single Municipality) through a hierarchical selection type and the variables of interest. As for the projections of population, it is possible to query the data under two different points of view: space or time. In the first case, for each year data are arranged by age and region. In the second case, for each region, data are arranged by age and year. For the life tables it is possible to choose territorial level, type and sex. The result of the query is displayed by the user's browser, but can also be saved directly in text format and therefore be used for personal elaboration with spreadsheets on the user's own personal computer.

DATA ANALYSIS

The Health for all (*HFA/ITALY)* database has been developed to fulfil the need of an information system capable to describe our country profile, in terms of use of health services and its resources, health conditions and life styles of the population under a geographical and temporal point of view. This system is integrated in the sense that it is an extended system with multiple thematic areas coherent both internally and among each other by using standard definitions, desegregations and statistical guides. The system currently consists of 10 thematic areas for a total of about 4,000 indicators available on provincial and regional level in a time series. All indicators referring to population are stratified both by sex and by age. Furthermore, data are standardised where bias due to different age structures can occur. The indicators are complemented with a statistical guide providing information on the indicator contents, the definitions used, the description of the calculation method, the list of data sources and all that is necessary to use the information correctly. Specific software has been used for information reference and analysis. It was originally implemented by the World Health Organisation (WHO) for the management of the "Health for All by 2000" plan indicators and subsequently modified by WHO experts adapting it to the necessities of ISTAT indicator system. It is a

very user-friendly inquiry and presentation software enabling the simultaneous display of different indicators, the building of tables, the creation of maps and graphs, and therefore a statistical analysis of the indicators included in the data base.

For more information on the above initiatives, please contact Alessandra Burgio (burgio@istat.it) or Marina Venturi (venturi@istat.it).

MADAGASCAR

from Julia Ravelosoa

Rural market survey in Madagascar. Economic poverty affects 77 percent of the Malagasy rural population, which is mostly agricultural. The increase and stabilization over the long term of the income of rural households are the two most important strategic objectives for reducing rural poverty in Madagascar. Although increasing agricultural production, both for self-sufficiency and for revenue-generation, along with other rural products is essential, it is important that 1) the prices of these products be lucrative and 2) there be a fair exchange between the urban environment (source of manufactured and imported products) and the rural environment (source of agricultural, livestock, fishing and forestry products).

The National Statistical Institute (INSTAT) conducted a "Rural Markets Survey" (RMS) to identify the elements for the development and implementation of a policy for the enhancement of rural market activities as a means of combating rural poverty. The initial survey was limited to the two provinces of Antananarivo (high plateaux with predominantly self-sufficiency production) and Toamasina (coastal region with predominantly cash crop production) and the observations refer to the "*agro-ecological zones*" in which "rural development task forces" (RDTF) are established.

Survey design: The approach used for the RMS2001 was that of purposive selection. The survey design takes into consideration the supply of the markets in the universe and the variability of prices associated with that distribution. In order to better capture the price dispersion, data collection was done at different points of sale where price levels depend on the level of consumer attendance. The universe consists of all rural markets in the two provinces of Antananarivo and Toamasina, thereby excluding urban markets. To develop the survey frame, the agro-ecological zones were considered as strata and within each stratum there were two sub-strata consisting of the "main rural market", the number of which has to allow for consideration of the variability of prices (2 or 3 main markets, as applicable) and the "secondary rural market" (2 per main market). Lastly, to ensure the representativeness of the sample units, rural markets were selected in each sub-stratum in each of the series of agro-ecological zones. Thus, the sample was composed of 45 rural markets (24 in Antananarivo and 21 in Toamasina), including 15 main markets.

*Choice of products and items***.** The list of products and items to monitor was developed on the basis of two major concerns: 1) the specificity of the production in the agro-ecological zones and that of the final consumption items in the rural environment; and 2) the comparability (spatial and temporal) of the products and items by ensuring that they had precise definitions and specifications. A *product* is a good that results from a production process. In the RMS2001, four categories of products were defined: 1) agricultural, livestock and fishing products; 2) manufactured products from industrial plants; 3) inputs, and 4) craft products. Essential products may be included in these various categories and are distinguished by the fact that they are indispensable, indeed vital, to the survival of the rural population. An *item* is a product with precise specifications with a view to commercialization and which meets the needs of clients.

Questionnaires and structures: Determining the spatial and temporal variability of prices is among the main objectives of the survey and is one of the major reasons for developing two types of questionnaires for the RMS.

(1) The "*community questionnaire*" (CQ) which collected data from elected officials, administrative authorities, associations, etc and contained five sections dealing with a) identification and follow-up of the survey, b) characteristics of the market, c) shipping of products, d) production and supply, and lastly e) price levels in 2000. Only one community questionnaire was completed per rural market. The data collected on product prices relates to the 2000 base year, which means the survey is retrospective.

(2) *The "price questionnaire*" (PQ) surveyed items in November 2001 at points of sale and contained five sections: a) survey of agricultural prices to producers, b) survey of the prices of livestock and fish products to producers, c) survey of the prices of craft products, e) survey of input prices to retailers, and lastly, f) survey of prices of household consumption products. The price questionnaire (PQ) was completed through in-person interviews with merchants in the market at the time of the survey. This means that each main rural market was visited four times (once per week on market day), while each of the secondary rural markets was visited twice (once every two weeks). The prices recorded were those of the day of the visit during the month of November 2001. *Three categories of prices* were collected during the survey: a) price to rural producers, b) price to rural consumers, and c) shipping cost. The *unit of measure for quantity* used in rural markets is not always the same as that of the international system. Enumerators were asked to record the units used in the communities when gathering the prices of items; conversion was done afterwards during data processing using the rates obtained in the field.

Survey schedule: The field survey lasted one month. It began simultaneously in all rural markets on October 29, 2001.

Survey personnel: The survey required 60 enumerators, 15 comptrollers and 6 zone chiefs under the direction of supervisors who were statistical technicians and technical officers.

Survey results: The survey results are in the process of being published.

For further information, please e-mail Randretsa Iarivony, Economic Statistics Branch, INSTAT Antananarivo, Madagascar, dginstat@wanadoo.mg.

PHILIPPINES

from Gervacio G. Selda, Jr.

The National Statistical Coordination Board (NSCB), through its Statistical Survey Review and Clearance System, has approved the conduct of the **2002 Information and Communications Technology (ICT) Survey in the Government** by the National Computer Center (NCC). While the National Statistics Office's ongoing Survey on ICT of Philippine Business and Industry covers the private sector, this NCC survey covers the government, thereby providing a complete picture of ICT in the country. The objectives of the survey are: (a) to support NCC's data requirements for the development, dissemination, and interpretation of policies and guidelines for Electronic Data Processing (EDP) applications, procurement and operations in the government; (b) to monitor ICT adoption and use in the government and come up with ICT trends; and (c) to maintain a databank on ICT statistics. The survey covers all national government agencies (NGAs) and their regional offices including attached bureaus and offices, government-owned and controlled corporations (GOCCs), state universities and colleges (SUCs) and local government units (LGUs). Two questionnaires will be used in the survey, entitled: (a) ICT Survey in the Government, and (b) LGU E-Governance Readiness Survey. The LGU survey is one of the components of the project, entitled "Jumpstarting Electronic Governance in Local Government Units" or "e-LGU" which aims to give LGUs a head start in computerization by making them more capable of embracing ICT in their operation for better public service. The survey will also help in choosing the 100 LGUs nationwide which will be the project beneficiaries for the first year. The major data items included in the questionnaire for NGAs, GOCCs and SUCs are e-commerce and networking, hardware/computer equipment, software and information systems, ICT manpower and organization, ICT budget and spending, ICT projects, and ICT practices and policies. For the LGU questionnaire, the major data items are hardware/equipment, network, Internet and e-commerce, investment and planning, application system, ICT organization and planning, and project commitment. The results of the survey will be released in January 2003.

The NSCB has also approved the conduct of the **Survey on Value Chain Analysis (VCA) on Holiday Décor** by the Bureau of Domestic Trade of the Department of Trade and Industry. The survey has the following objectives: (a) to determine the flow of holiday decor, particularly Christmas balls, from the source of its raw materials until it reaches its final user, the consumer; (b) to identify the value-adding inputs and economic activities that this commodity passes through; and (c) to measure the value added as the commodity flows through

the chain considering the supply process, production process and marketing activities and determine how these affect the price structure. The data will be gathered through the conduct of interviews with the players of every economic activity or channel in the chain that the commodity passes through starting at the source of raw materials. Respondents will come from the holiday decor's top exporters and manufacturers for local sales as well as major suppliers, distributors and marketing firms. Three questionnaires will be used in the survey: (a) for the supplier of raw material, (b) for the manufacturer, and (c) for the distributor. The major data items in the questionnaires include organizational background, product design, procurement, materials quality control, production, postproduction quality control, packaging and labeling, pricing, promotion, sales, and problem areas. The survey results are expected to be available in 2003.

Survey clearance was also granted by the NSCB to the **2002 General Survey on Labor Organizations** of the Bureau of Labor and Employment Statistics under the Department of Labor and Employment. The survey is undertaken every four or five years to generate updated information on federations/labor centers. The results of the survey will serve as basis for the assessment of the present state of trade unionism and collective bargaining in the country. It will also be used in identifying areas in need of government assistance and intervention to further strengthen the federations' machinery towards the promotion of wider industrial peace and harmony. The survey covers all registered federations/labor centers in the country, numbering to around 177. The data items to be collected are number of affiliates, membership, collective bargaining agreements, program/activities conducted, industrial action, conventions and publications, and issues and concerns. The results of this survey will be released in March 2003.

The NSCB Executive Board, the policy-making body in statistics here in the Philippines recently approved the recommendation of the Inter-Agency Committee on Labor, Income and Productivity Statistics and the NSCB Technical Staff for the **computation of provincial poverty thresholds and incidence** including poverty gaps, income gaps, and the Foster Greer Thorbecke indices in accordance with the poverty estimation methodology developed by the NSCB. The proposed methodology was based on the results of the project, "Strengthening Institutional Mechanisms for the Convergence of Poverty Alleviation Efforts", with financial assistance from the United Nations Development Program for the development of the methodology for provincial poverty estimation. The Food and Nutrition Research Institute developed four types of menus for each province from which a particular menu was selected as basis for the estimation of the food threshold taking into consideration the minimum energy, protein, and other nutrient requirements, food commonly eaten and the generally low cost commodities available in each province. Such methodology has been validated among planners and decision makers at the regional levels and in the local government units, local nutritionists, and other stakeholders of poverty statistics nationwide in a series of users' fora conducted in Metro Manila, Baguio City, Tacloban City, Iloilo City and Davao City in CY 2000. There is adequate data support for the computation of provincial poverty statistics based on this methodology.

Also, the NSCB Executive Board approved the conduct of **"Performance Measurement Scheme for National Statistical Agencies"**. The primary objective of this undertaking is to formulate a mechanism for monitoring the performance of statistical offices and other data producers while the specific objectives will focus on the development of a framework and the formulation of an evaluation system including a list of indicators to measure and assess the performance of statistical agencies. The performance of an office shall be measured and assessed on the following areas of concern: (a) agency's statistical outputs/products/ services vis-à-vis agency mandate; (b) manpower resources; (c) financial resources; and (d) equipment/facilities. This activity will initially cover nine government agencies that agreed to take part in the exercise. The results of monitoring are very useful for advocacy, promoting public accountability of statistical offices and rationalizing allocation of resources. Monitoring of performance is not envisioned to compare the different statistical agencies because each has unique functions and varying activities but the results shall be used as benchmark information for regular updating of statistical standards. For more details on the NSCB reports, contact: Dr. Romulo A. Virola, Secretary General of NSCB at www.nscb.gov.ph.

The Bureau of Agricultural Statistics (BAS) is currently conducting a study entitled **"Development of Integrated Agricultural Survey (IAS) Design: An Alternative to BAS' Specialized Surveys"**. The study was conceived to gather reliable and timely information on crops and livestock production and other aspects of agriculture through a unified and integrated survey. It made use of information from the 1999 Barangay Screening Survey and the 2000 Household Screening Survey to come up with the sampling frame and sampling units for the proposed survey, respectively. Based on the results of the study, the proposed survey exhibited the advantages as well as the challenges of integrating agricultural commodities into one survey. The efficiency and practicality of estimates varied among different commodities and among different indicators. There was no specific pattern as to what is the best approach to a particular estimation procedure. In cases where design-unbiased estimation cannot provide the required efficiency level, alternative estimation procedure can be applied. IAS will surely minimize the cost of survey operations, though estimation procedure differs according to indicators and commodities. For more details, contact: Mr. Romeo S. Recide, Director of BAS at RSRecide@mozcom.com.

The Statistical Research and Training Center (SRTC) in cooperation with the World Bank Institute and the Philippine Institute of Development Studies recently conducted the "**Distance Learning Course on Basic Poverty Measurement and Diagnostics"**. The training was held on June 10-22, 2002 at the AIM Training Center in Makati City, Philippines. The 2-week course featured a daily two-hour multipoint videoconference session with France, Thailand, Indonesia, Vietnam and Sri Lanka, face-to-face lectures with selected local experts/trainors on poverty concepts and the use of STATA for poverty analysis. Twenty-five participants from various government agencies and institutions involved in poverty monitoring and analysis in the country attended this training.

Also, the SRTC recently completed two of its in-house researches under the Re-engineering the Government Statistical Services Project – Phase II. The first one is entitled **"A Comparative Study of Seasonal Adjustment Methods for Philippine Time Series"**. The study explored the applications of X12 RegArima and TRAMO-SEATS deseasonalization approaches to some Philippine time series data using a new freeware from EUROSTAT called DEMETRA, which interfaced these two seasonal adjustment methods. Theoretically, explicit model based procedures, such as TRAMO SEATS, ought to be better than the X-11/X-12 family since the latter's filters do not take into account the statistical properties of the time series while the TRAMO-SEATS filters are mean-squared error optimal. This research undertaking sought to consider, on the grounds of some empirical criteria, which procedure ought to be preferred for some Philippine time series. For the domain of Philippine time series studies, TRAMO-SEATS was recommended. The other completed research is entitled **"Constructing Self-Organizing Poverty Maps from the FIES and APIS"**. The study exhibited an alternative mapping system that visually displays the structures of welfare relations among comparable geographic areas at the sub-national level (such as regions and provinces) based on a poverty database sourced from the Family Income and Expenditure Survey (FIES) and the Annual Poverty Indicator Survey (APIS). This mapping system is based on the Self-Organizing Map (SOM), also called the Kohonen Map or the Self Organizing Feature Map. The results suggest that poverty structures and welfare relations among the regions do not change considerably over a period of time of three years. Moreover, poverty maps indicate that poverty in the Philippines is largely a rural phenomenon. Poverty reduction programs that aim to enhance income-generating capabilities must be targeted toward the rural areas. For more details on SRTC studies, contact: Mr. Gervacio G. Selda, Jr., Executive Director of SRTC at srtcoed@srtc.gov.ph.

POLAND

from Janusz Wywial

Polish household surveys are in a transition period. This is mainly connected with administrative transformation. Moreover, standards of survey sampling used in the European Union are being adapted to Polish circumstances. New sampling designs are being proposed or existing ones are being redefined. The population is stratified in a new way. Optimal sample sizes are determined anew. These problems deal with surveys carried out by the Central Statistical Office in the last decade, that is, during the transition

period. More details can be obtained from J.Kordos@stat.gov.pl, B.Lendnicki@stat.gov.pl and M.Zyra@stat.gov.pl.

The conference "90 Years of Polish Statistical Association" was held in Modlnica near Cracow on July 14-15, 2002. The proceedings of the Conference were connected with the following topics: Statistics in XXI century, ethical problems in statistics, statistics and informatics, and quality of statistical data. More details can be had from S. Smiech, smiechs@janek.ae.krakow.pl.

The conference "Survey Sampling in Economical and Social Research" was held April 23-24, 2002 in Katowice, Poland. It was organised by the Department of Statistics of the University of Economics in Katowice and the Department of Statistical Methods of the University of Lodz. The main topics of the Conference were as follows: Problems connected with design effect, testing of statistical hypothesis in the case of complex samples, application of some discrimination methods to mean value estimation in the presence of nonresponse, internet statistical surveys, survey sampling in auditing problems, small area estimation, reliability of stock data, survey sampling on more than one occasion, optimisation of sample size in the case of estimation of more than one mean value, estimation of variances of statistic from complex samples, simultaneous stratification of population and optimal sample allocation in strata, estimation of quantiles, prediction of vector of total values, classification estimators of mean value supported by auxiliary variables. More information is available form J. Wywial, wywial@ae.katowice.pl or www.ae.katowice.pl. The next conference will be held November 4-6, 2003 in Lodz, Poland. More details: wywial@ae. katowice.pl.

SPAIN

from Montserrat Herrador

Co-ordination of business samples using the Permanent Random Number Technique at the Spanish National Statistics Office (INE). Public administrations are requiring more and more information from enterprises, and particularly National Statistical Institutes (NSIs) for statistical surveys, which results in heavy workloads on enterprises. In order to control the response burden on enterprises, INE has implemented techniques of negative co-ordination in sample selection.

The use of these techniques has been made possible by the existence of the following tools within INE:

- Central Directory of Enterprises (DIRCE) which is the single sampling frame for use in business surveys.
- System of monitoring collaboration of the reporting units (DICOIN) which traces at the microdata level indicators for each enterprise selected to participate in some of the economic surveys performed by INE. Several analyses can be derived on the number and type of questionnaires sent and received by the sampling units, collaboration attitude, contacts, and non-response.

From the year 2002 and due to the increasing number of economic surveys, the Permanent Random Number (PRN) technique is being applied in order to reduce the overlap between enterprise samples for different surveys (negative co-ordination). A major feature of this kind of technique is its ability to negatively co-ordinate several surveys in a simple fashion.

The PRN technique for simple random sample without replacement (SRSWOR), introduced by Atmer et al. (1975) at Statistics Sweden, provides a solution to reduce the overlaps between samples across surveys and over time.

The following algorithm is applied for selecting a SRSWOR: each unit in the list frame is assigned a random number drawn independently from the uniform distribution on the interval [0,1]. Let X_i denote the random number assigned to unit *i*. The frame units are sorted in ascending order of the X_i. A sample of size *n* is composed of the first *n* units in the ordered list.

Depending on both the order of random numbers in the list frame and the origin of the sample selection (the starting point to select the units which are to be included in the sample), samples are obtained with varying degrees of overlap.

The success of the negative co-ordination will depend, on the one hand, on the number of sample and population units in each stratum and, on the other, on the similarities among strata from different business surveys. For these surveys, generally speaking, population is stratified according to very

similar variables (geographical code, economic activity, enterprise size, etc.) and this fact permits a satisfactory negative co-ordination.

Each unit from the list frame has related a number, response burden number (RBN), that measures the burden response over time. RBN is calculated using the information from DICOIN. Initially RBN was set equal to 0 for all units. When the unit is selected for a structural survey, the value 1 is added and when it is selected for a short time survey, the value 0.5. So, this number measures the cumulated response burden of the unit.

To avoid the case that a newcomer has a higher probability in next samples, an artificial RBN is assigned. When new units are added to the frame, random numbers are generated. These random numbers should start at the place in the cycle where the latest assignment stopped. The value that takes the newcomer as response burden number is the RBN of the previous unit according to the order of random numbers.

The historical information about the collaboration of each enterprise with INE business surveys also leads to control which units are selected for the sample because of its very small population stratum size. In these cases, the PRN technique is used for sample rotation and if it is possible, to guarantee that a unit will not belong to a survey sample for more than five years. For more information, please contact Montserrat Herrador, herrador@ine.es.

UNITED KINGDOM

from Peter Lynn

An important new longitudinal survey began data collection this year. Wave 1 of the **English Longitudinal Study of Ageing** (ELSA) has been completed. ELSA aims to become an interdisciplinary data resource on health, economic position and quality of life as people age. It covers the broad set of topics relevant to a full understanding of the ageing process. ELSA is beginning life with a cross-sectional sample of persons aged 50 or over, who will be followed up regularly until death. This sample will be supplemented with new cohorts on a regular basis (e.g., 50-53 year-olds every four years), so that the survey will eventually become a series of age cohorts. The survey is initially funded by the US National Institute of Ageing and a consortium of eight UK government departments and is being carried out by a multi-disciplinary team from the Institute for Fiscal Studies, the National Centre for Social Research and University College London. See: http://www.ifs.org.uk/elsa.

The Department for Education and Skills is developing another new longitudinal survey towards the other end of the age spectrum this time. The **Longitudinal Survey of Young People in England** (LSYPE) will interview young people annually from approximately age 13 until age 23 with a starting sample of at least 12,000. The survey will focus primarily, but not exclusively, on education, skills and employment issues. A range of feasibility and development projects has been commissioned in 2002 and will continue into 2003. Main stage fieldwork is scheduled to begin in Spring 2004.

A new **Survey of Self-Reported Offending** behaviour is being planned by the Home Office. A number of preparatory and feasibility studies have been carried out, including both desk reviews and field pilots. These are available, along with the latest news on plans for the new survey, at http://www.homeoffice.gov.uk/rds/offending1.html.

The **Health Survey for England** (HSE) is a large survey with continuous fieldwork, involving CAPI interviews and a nurse visit to take various measurements. The survey, which began in 1991, has a constant core module, plus a substantive and/or population focus that changes from year to year. The focus in 2003 will be cardio-vascular disease and in 2004 it will be ethnic minorities. For the latter, a boost sample of non-whites will be selected and questionnaire topics of particular relevance to minorities will be included. This will be the second time that ethnic minorities have been the focus of the survey—the first having been in 1999. Further details of the survey, including copies of published survey reports, can be found on the HSE website: http://www.doh.gov.uk/public/hthsurep.htm.

Results from the **Northern Ireland Health and Social Well-Being Survey** have been published. The survey was carried out over six months in 2001, with a final sample size of 5,205 interviewed adults. Topics covered include general health, smoking and drinking, illnesses, sexual and mental

health. Interviews were carried out face-to-face, with a self-completion instrument for the more sensitive questions. Findings and methodological details can be found at http://www.nisra.gov.uk/whatsnew/wellbeing/index.html.

A new **Integrated Social Survey** is being planned by the Office for National Statistics. This would combine four major current surveys—the Labour Force Survey, Expenditure and Food Survey, General Household Survey and ONS Omnibus Survey—into a single operation. Sample design, field work and estimation methods would all be integrated. A simple random sample of addresses is envisaged, with in-home face-to-face interviews. Advantages would include increased precision of estimation for all the surveys, improved inter-censal monitoring of key social and socio-economic measures such as housing, employment, ethnicity, education and health, and provision of certain small area statistics. Contacts: Nikki.Bennett@ons.gsi.gov.uk or Tony.Manners@ons.gsi.gov.uk.

Since 1967 the **English House Condition Survey (EHCS)** has run every five years, measuring the state and condition of the housing stock. From April 2002, it has moved to a continuous format. This will enable monitoring of progress towards the new targets set by the Government for the delivery of decent social housing. The sample design has also been modified, following a feasibility study, so that rare tenure groups are now identified using another survey, the Survey of English Housing, as a screening instrument, whereby respondents to that survey are asked about the tenure of a randomly-selected nearby address. The EHCS has four components—a household interview, a visit by a professional surveyor to assess the condition of the property and take digital photographs, a "market value" survey, in which local valuers assess the market value of each property based upon the photographs and other survey information, and an interview with the landlords of privately renting tenants identified in the interview survey. The re-designed survey will deliver an annual sample of some 8,000 interviews with an associated physical survey. Fieldwork will run in 4 seasonal blocks each year starting in April, July, October and January. Contact John.Flatley@ons.gsi.gov.uk.

To All Members

- The IASS needs your contribution.
- Please do not forget to renew your membership.
- As of January 2002, French Francs are no longer accepted. As a consequence, the payment of dues and subscriptions must be made in either Euros or U.S. dollars.

Strategic Objectives of China's National Statistics System

Zhu Zhixin
Commissioner, National Bureau of Statistics of China

I. The Significance of Developing Strategic Objectives for China's National Statistics

China's statistics system has been developing since China adopted the reform and opening up policies more than two decades ago. This development has followed a course of resolute reform, persistent exploration, bold practice, and continued progress, and has made great contributions to China's socialist modernization drive. In the new century, China has entered into a new phase of building a nation with a comfortable society and accelerating the development of the socialist market economy. China's entry into the World Trade Organization marks a new era of China's opening up and reform practice. This has created new challenges for China's statistics system. Significant political and economic changes have been taking place everyday in the world. Science and technology have played a greater role in the social progress to an unprecedented degree. With the increasing demand for statistics, new theories and methodologies in statistics have occurred, thereby enriching statistical practices. The need for statistics has created great opportunities and challenges for China's statisticians.

The strategic objective for a national statistics system has overall long-term and fundamental impacts on the development of the nation's statistics. The objective is aimed at promoting the development of China's statistical reform, having the national statistics system play a full role, meeting the needs of the socialist market economy, and guiding the development of China's statistics system in the new century. Defining the strategic objective will help the Chinese statistical offices and statisticians reach a common understanding and coordinate their actions in order to deal properly with relations between the whole and the parts and between long-term and immediate interests. It will also make the best use of the human, financial, material, and technical resources of the national statistical system, accelerate the modernization process of statistics, and provide quality services to the government and the public.

II. Strategic Objective of China's National Statistics

For the coming period, the strategic objective of China's national statistics system should be to build a modern national statistical system characterized by "hi-tech facilities, quality statisticians, efficient mechanisms, and quality information." This system should be able to meet the requirements for the socialist market economy while remaining in line with China's particular conditions and international standards.

The socialist market economy system determines the main framework of the statistical system. The statistical system cannot surpass nor lag far behind the development of the economic system. The statistical system serves the socialist market economic system. Its success will be measured by how well it meets the demand of the socialist market economic system.

Adaptability is a basic requirement for developing the strategic objective of the national statistics system. Adaptability means that the strategic objective must start with China's basic situation and the reality of China's statistical work, build on the lessons learned from our past statistical reforms, and make a good analysis of the future development trend.

Conformity to international standards is another basic requirement for developing the strategic objectives. With the economic globalization process and China's entry into the World Trade Organization, the Chinese economy is becoming more and more integrated into the world economy. There are ever-increasing demands for internationally comparable statistics. International organizations and the entire statistical community have set up international standards for use. The accumulated experience and the technology applied in the area of statistics are extremely beneficial to China. The establishment of a modern national statistical system in conformity with international standards will facilitate further reforms and opening up of the Chinese economy. The modern system will also enhance the economic,

social, and cultural exchanges between China and other parts of the world.

"Hi-tech statistical facilities" refers to the statistical theories, methodology, and informatics that the national statistical system employs in the statistical process and in all the aspects of its work. The statistical process covers program design, data collection, data entry, checking, transmission, processing, storing, management, analysis, and dissemination. The aspects of the process include statistical research, education and training, publishing and printing, office automation, financial management, and logistics. The beginning of the 21st century has been called the hi-tech era and the information era. The development of information technology has greatly impacted every phase and aspect of statistics. The entire statistical system must adopt the concept of "hi-tech" and make full use of scientific statistical methodologies and advanced information technology. This is a prerequisite to enhancing the efficiency and quality of statistics. Statistics cannot deal with the challenges of the new century without the support of the hi-tech measures.

"High quality professionals" refers to the fact that the national statistical system must have a large number of professionals with statistical, economic, or other backgrounds, who are well-educated in profession, rich in experience, creative in their work, dedicated to their profession, and high in professional morale. Well-qualified professionals are the basis for ensuring the application of advanced statistical theories and methodology, the use of hi-tech facilities in statistics, the improvement of statistical efficiency, the quality of statistical information and the development of statistics in the new century. The modern statistical system cannot be established without these high quality professionals.

Efficient statistical mechanisms include an efficient management system, a legal system, and financial support mechanisms. The past has shown that a rational management system, powerful rule-of-law conditions, and sufficient financial inputs are necessary in order to adopt scientific statistical theories and methodologies, to apply advanced information technology for statistical purposes, to attract, train, and use quality statistical professionals, and to prevent various kinds of interferences in order to assure data quality.

Quality statistical information is defined in terms of "rapidity, fineness, and accuracy" of the information produced by the national statistical system. "Rapidity" is the requirement for timely availability of statistical data, "accuracy" refers to the objectiveness of the information, and "fineness" means the relevance, systematization, comparability and consistency of the products. Quality is the life of statistics. Therefore, ensuring quality is the core of the national statistical strategy.

The relationship among the hi-tech facilities, quality professionals, efficient mechanisms, and quality information is clear. Hi-tech facilities are a prerequisite, quality professionals are the basis, efficient mechanisms are the key, and quality information is the goal. These four aspects are very much interrelated and complementary to one another and represent the four major features of a modern national statistical system.

III. Key Strategic Areas of China's National Statistics

To reach the strategic development objectives for national statistics in the new century, we should set the following six major areas as priorities of development.

1. Establish a national accounts system that is compatible with international standards and conforms to the needs of sustained economic development

China's new national accounts system is established following the concepts, principles, methodology, and framework of the 1993 United Nations System of National Accounts (SNA). However, there are still major differences between the two. Such differences exist because China has long practiced the Material Product System (MPS) and because the basis of the national accounts is weak. China should enforce its current system to comply with the SNA while removing the remaining components left over from the MPS, and then make some adjustments to the accounts and basic accounting sheets. In view of the trend of social development in the future, the main factors that will restrict economic development will be excessive population growth, lack of natural resources, and deterioration of the environment. Therefore, it is a long-term and hard task to reflect the needs of sustainable development and measure the damage to the ecology, the deterioration of the environment, and the decrease of resources due to social development and economic growth.

2. Establish a scientific methodology system for the national statistics system

A scientific methodology system that conforms with the socialist market economy must be established.

Currently, the following problems remain in China's system: (a) the statistical standards are not yet complete and unified and the current standards are not yet well-implemented; (b) the system of statistical indicators is not systematic, coordinated, and standardized with the coexistence of excess and shortage; and (c) methodologies of statistical surveys cannot meet the requirements, data quality from statistical reports is not ensured, sample survey methodology cannot meet the needs of the conduct of multiphase and multipurpose surveys, censuses have become a heavy burden, and data from different sources are not consistent with one another.

The solution to these problems is three-fold. First, China must establish and implement a unified national system of statistical standards that conforms to the international ones and adapts itself to China's situation. Second, it must establish a systematic, standard, and easy-to-use system of statistical indicators that meets the needs of the socialist market economy and sustains development of the national accounts system. Third, China must set forth an advanced, suitable and reasonable methodology system that has periodical censuses as its basis, current sample surveys as its main methodology, and other kinds of surveys as its supplements.

3. Push forward all-around informatics development and establish a national statistical information network system

Networks characterize the information era. Because there is a fierce competition in the area of information capacity, the construction of a national statistical information network system is of great importance.

The basis of the national statistical information network system is the national statistical information databank. The databank refers to a dynamic process in which the national statistical system gathers abundant raw data (figures, texts, and pictures) and has the data edited, processed, developed, utilized, and maintained. The system should put special emphasis on the efficient use of data resources and a complete solution to the problems that now exist with the statistical information. For example, although much data has been collected, very little has been processed and tabulated. In other cases, much data has been isolated, but little has been integrated and shared. Whereas much of the data is in traditional format, very little exists in electronic format. Much of the current information is static, and not dynamic.

Given the current situation, we must attempt to build a comprehensive, practical, and open national information network system that is based on the internet, supported with advanced technologies, guaranteed with unified and standard management, and composed of the databank as its principal body. The system should efficiently provide comprehensive and in-depth products to meet the requirements of all categories of users. Eventually, it should support the macro economic control and provide services for the public. Information technology should be used in all the phases of the statistical process including collection, transmission, processing, storage, analysis, and dissemination so as to improve the efficiency, quality, and information handling capacity of the national statistical system.

4. Develop statistics with science and training, establish the mechanisms of building initiative and human resources development, and develop a large contingent of quality statistical professionals

"Promoting China's development with science and technology" is one of China's basic development strategies. Emphasis must be placed on taking initiative to develop science and technology and statistical education and training if China is to solve its problems in the statistics area and reach its development goals. Steps must be taken to speed up the building of a vigorous mechanism of science and technology initiative, encourage more statistical academic exchanges with other nations, explore the theories and methodologies that are of great importance to our statistical practice, encourage the prompt and effective use of statistical research results in statistical practice, and make full use of statistical science and technology initiatives in the modern statistical system. The statistical system must make its best effort to develop its human resources, create competition and inspiration among its staff, and bring forth a large talent in statistics. That talent should include professionals in four areas: objective and effective information collectors, unified and coordinated statistical designers, capable and active statistical analysts, and ethical and fair-minded statistical managers and law executors.

5. Enforce the statistical legislation, reform the statistical management system, establish a mechanism to ensure the financial resources, and realize the objectives of the statistical strategy

Since China started its reform and opening up policies, statistical legislation has been enforced gradually. However, two major problems exist: (1) the statistical legislation is not yet complete and cannot meet the requirements for the statistical development under the rule of law, and (2) violations often take place and cannot be stopped effectively. Under such circumstances, we must take the law enforcement as a long-term and basic strategic measure to ensure data quality. In order to manage statistics under the rule of law, we shall establish a strict and powerful statistical legal system. This will enable all individual links and areas of statistics to base themselves on the law and any violations to be investigated.

To a large extent, the statistical management system determines the orientation of statistical services, the efficient use of statistical resources, and the realization of the functions of statistics. Several problems currently exist within China's statistical system: (1) the system cannot effectively resist interventions; (2) the relationship between the national bureau and local bureau is difficult; (3) the programs of the national office and ministries are not well-coordinated; (4) the division of work is not clearly defined; (5) the responsibilities are duplicated; and (6) the management is not standardized. The management structure has greatly affected current statistical reforms and data quality. Reforming the national statistical management system should be accomplished by the gradual establishment of a system with centralized management, with the responsibilities of the national bureau, local offices, and ministerial offices clearly defined and coordinated, and with the entire statistical system functioning effectively.

The conduct of statistical programs and the improvement of data quality cannot be made without necessary conditions, especially financial, being met. The main problem now is the serious insufficiency of financial resources. Solving this pressing problem will require us to actively explore new sources of supply, curb unnecessary spending, and utilize all assets. A reliable and efficient financial supporting mechanism shall be established to initiate statistical programs, associate the programs with financial resources, and ensure the smooth conduct of the programs.

6. Establish a powerful national statistical information management system in order to ensure data quality

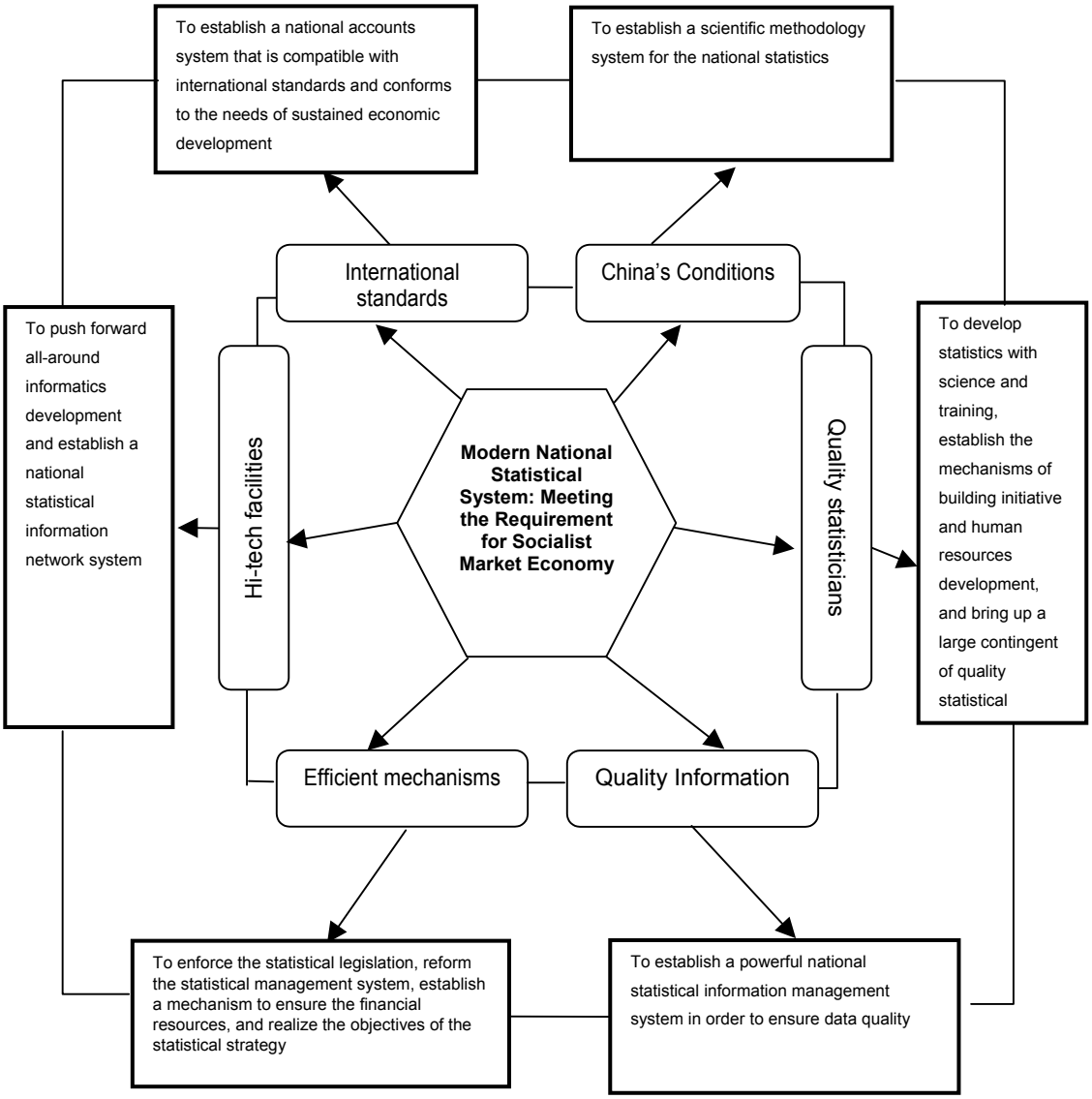
It is a core task and basic responsibility for the national statistical system to provide quality data. We must establish a powerful national statistical information management system that controls all the stages of the statistical process, all the areas of statistics, and all the individual statisticians so as to ensure the accuracy and timeliness of the statistics to be released. Quality control shall be carried out in every stage, in every area, and by every person. China's participation in the General Data Dissemination System (GDDS), developed by the International Monetary Fund (IMF), has demonstrated its efforts in this regard.

Looking ahead, China's statistics system will have a very bright future. Establishing a modern national statistical system will greatly push forward the development of China's statistics. The extensive use of hi-tech measures will become the driving force for this development, for which quality human resources will lay a solid foundation. These effective mechanisms will invigorate China's statistics, thereby creating quality statistical information that will help the national statistical system earn its credibility.

We would like to stress that during the implementation of the national statistical strategy we shall take the responsibility of providing quality data services to the government and to the general public. We shall have a greater understanding as we learn from applying the advanced theories, methodology, and skills that are the essential conditions for further development of statistics. We shall have a greater sense of initiative in conception, methodology, measures, organizational structure and theories, all of which are the forces of statistical development. We shall have a greater sense of coordination, which will help organize and manage the limited resources and make joint efforts to develop statistical functions. We shall have a greater knowledge of the laws of statistics, thus allowing us to penalize violations of the Statistics Law in order to ensure data quality.

As the new century begins, we are full of confidence and expectations. As long as we strengthen our management, unite, take the correct directions, make continued efforts, and keep progressing, we are sure we will reach our objectives and the modern national statistical system will make greater contributions to China's modernization process.

Framework of the National Statistical Strategic Objectives



Sample Design of the Survey on Population Changes and Family Planning in Vietnam, 2001-2005

Nguyen Quoc Anh
Director CPSI, Vietnam

1. Objectives and Requirements

Implementing the interministerial circular No. 02/2001/TTLT/TCTK-UBQGDS on 14 February 2001 between the General Statistical Office (GSO) and the National Committee for Population and Family Planning (NCPFP), on 23 February 2001, the GSO promulgated the Plan of Survey No. 95/TCTK-DSLD. According to the plan, the survey on population changes and family planning would be carried out on the first day of April every year in the 2001-2005 period in order to estimate data of births, deaths, in- and out-migration of population, and the contraceptive use of currently married women of reproductive ages (15-49) at both the provincial and city level.

The sample design adheres to the following requirements:

- The sample selected must be representative for province, city, and region, both urban and rural, of the entire country.
- Sample selection shall be based on available documents of the 1 April 1999 Population and Housing Census such as lists of communes and wards, lists of enumeration sites, maps and lists of enumeration sites for selection of sample and preparation work of the survey.

2. Coverage and Sample Design

2.1 Coverage

In order that the survey sample is representative for provincial, city, and regional levels, and urban and rural areas, the determination of sample size is based on the number of events needed to collect data on births and deaths; at the same time it must ensure the comparability of results between provinces and cities. From 2001 to 2005, each province and city must carry out a yearly survey with a sample size of about 24,000 inhabitants; it is equivalent to 5,500-5,800 households. Based on the size of each province and city, the Department of Population and Labour will determine the number of enumeration sites needed for all provinces and cities by urban and rural areas.

2.2 Sample Design

The following elements describe the sample design:

- The basic sample unit is the enumeration site with about 100 households ($\pm$20 households).
- The sample frame is based on lists of enumeration sites established in the 1999 Population and Housing Census according to two areas, urban and rural sites, in each province and city. There are thus 122 sample frames in the whole country.
- For sample allocation, the sample population of each province and city is divided into two subgroups, urban and rural, based on a percentage of the urban/rural population of each province and city. Then the number of sample sites by urban and rural areas is defined.
- The method of selection used is systematic random cluster sampling with equal sampling probability of selection in each urban and rural area, namely:
  - The total number of enumeration sites of urban (or rural) areas of each province and city is divided by the number of enumeration sites to be selected in each area to calculate the skip interval used in the selection (as denoted by *k*).
  - A random number is selected to start (i.e., x: a number) ($x \prec k$). Succeeding site numbers are identified by the formula: $x_i = x + ik$ in which i = 1,2,3… and the series is stopped when all sample enumeration sites are identified.

Each province repeats the procedures separately for selection of rural and urban sites.

For urban areas:

- The total number of enumeration sites of the urban area of the province *i* is denoted by D_i^{TT}

- The total number of sample enumeration sites of the province *i* is denoted by m_i^{TT}
- Skip interval = $\frac{D_i^{TT}}{m_i^{TT}}$ (taking one decimal number).

For rural area:

- The total number of enumeration sites of the rural area of the province *i* is denoted by D_i^{NT}
- The total number of sample enumeration sites of the province *i* is denoted by m_i^{NT}
- Skip interval = $\frac{D_i^{NT}}{m_i^{NT}}$ (taking one decimal number).

As an example, for the rural area of An Giang province, there are 3,338 enumeration sites. The number of sample enumeration sites to be selected is 39.

- Skip interval is 3,338/39 = 85.6
- If, for example, the first random number selected is 8, succeeding random numbers to be selected are then 93.6; 179.2; 264.8; 350.4; 436; 521.6; … (the series is stopped when 39 numbers are identified)
- Thus enumeration sites with following codes are selected for the survey: 8; 94; 179; 265; 350; 436; 522; etc.

3. Steps of Selection

The Department of Statistical Population and Labour calculates and allocates the number of sample enumeration sites of each province and city by urban and rural areas and selects directly the number of sample enumeration sites by systematic random method (as noted above) to carry out the survey.

The lists and codes of enumeration sites selected and all copies of maps, listings of houses, households, and numbers of persons of each site are prepared by Department of Statistical Population and Labour and transferred to provinces and cities at training courses on the survey.

The statistical offices of the provinces and cities inform the statistical bureaus of districts of the lists of communes, wards, and towns (where there are sample enumeration sites) before the training of communes and wards. The statistical bureaus of districts coordinate with selected communes and wards to guide enumerators to check sites, update changing events (births, deaths, in- and out-migration) and make new listings of households to be surveyed (as the form established). Codes applied to enumeration sites are the codes used in the 1 April 1999 Population and Housing Census. Codes of administrative units (province/city; district; commune/ward) where there are sample enumeration sites are numbered according to the Viet Nam List of Administrative Divisions 2000 established by the Statistical Standards and Methodology Department, General Statistical Office.

4. Weights of Sample Data

Coverage of the sample survey of each province and city is nearly equal, but in practice, the proportions of standard population of some provinces are rather different, especially provinces just split (Kon Tum, Bac Kan). Therefore, the sample data are weighted following standard population structure in order that data collected from the sample reflect exactly the demographic characteristics of the population studied.

From the sample design, data of sample survey are weighted following average site size, standard urban/rural percentages of each province or city, and standard population percentage of each province, city in the whole country.

Processing data of provincial, city level:

Data of sample enumeration sites are multiplied by weights

$$w_{tinh} = w_1 \times w_2$$

where:

w_1 = weight based on average population size of enumeration site, calculated separately for rural and urban areas.

$$w_1 = \frac{\text{Size of standard site (or average number of inhabitants of a site)}}{\text{Size of each sample enumeration site (number of inhabitants of each site)}}$$

w_2 = weight based on standard percentage of urban/rural of each province, city (following official report on average population 2000, 2001, 2002).

$$w_2 = \frac{\text{Standard percentage of urban or rural population (complete data)}}{\text{Percentage of urban or rural population from sample survey}}$$

Processing data for eight geographic-economic regions and the whole country

Data of sample enumeration sites are multiplied by weights

$$w_{Toanquoc} = w_1 \times w_2 \times w_3$$

where:

w_3 = weight based on standard population percentage of each province, city in the whole country (following official report on average population 2000, 2001, 2002)

$$w_3 = \frac{\text{Percentage of population (complete data) of each province (as compared to the population of the whole country)}}{\text{Percentage of population from sample survey of each province (as compared to total sample population of the whole country)}}$$

Special Articles: Censuses Conducted Around the World

2002 Population and Housing Census in Latvia

The 2000 Population and Housing Census was the first census of population in Latvia since it regained its independence. The Census' aim was to obtain all-around and unbiased information on the number, composition, occupation, and sources of income of the population, as well as about their dwellings and other indicators included in the census program in each city and pagasts (the smallest administrative unit in the rural area) at a time when economic situation, population structure, and demographic processes have drastically changed. The necessity to carry out the census was also caused by the fact that 11 years had elapsed since the previous population census was conducted in 1989.

Preparation work to conduct the census was started immediately after the Cabinet of Ministers (CM) issued its regulations "On Preparatory Work for the Regular Population Census," under which the Central Statistical Bureau (CSB) was commissioned to prepare and conduct the regular census of the population. The Regulations dated May 11, 1999 of the Cabinet of Ministers specified the census date of March 31, 2000 and indicated that all information was to be collected and compiled pursuant to the situation on that date.

On December 16, 1999, the Saeima (Parliament of the Republic of Latvia) adopted the Law on Population Census, which provides a strong legal basis for the 2000 Population Census. Pursuant to the recommendations of international organizations and the experience of other European countries, the law envisaged that all resident population—natural persons registered in the Residents' Register, as well as those natural persons who are not, but have to be registered in the Residents' Register—has to be recorded in the census. The Law also requires that information characterizing the dwellings of these persons must be collected and compiled in the Census.

In compliance with the task set by the CM, a Government Commission was established to administer the preparatory work for the 2000 Population Census. The Commission solved different issues as to the preparation of the Census 2000 and supported the CSB in organizing and conducting the most important statistical activity in 2000.

The methodology of the census and the range of questions included were developed in compliance with the recommendations of the international organizations—UN Economic Commission for Europe (ECE) and "Eurostat," the Statistical Office of the European Communities—as well as the requirements of the state, local government, and other data users.

Two pilot censuses were conducted—one in two towns and two rural municipalities of Latvia in 1997 and one in Riga in 1999. They were conducted to verify different methods of data collection, the comprehensibility of the census questions, and the responsiveness of the population, and to test the data processing equipment, machinery, and software of the potential census data. Each pilot census involved approximately 5,000 persons.

Before the census started on March 31, 2000, the machinery necessary for data collecting and processing was gathered and the corresponding software was purchased. It involved scanners for questionnaires, computers, and and other equipment.

Information was collected by specially chosen and trained enumerators who visited the population at their place of residence from March 31 to April 30, 2000. If for some reason an enumerator did not meet with any of the inhabitants of a household, the residents could complete the census at one of the statistical offices or at one of 34 additional enumeration sections in Riga.

The critical census moment was set for Friday, March 31, 2000, at 12:01 a.m. Enumeration questionnaires were completed corresponding to the situation at that date and hour.

The 2000 Population Census recorded persons who were registered in the Residents' Register of the Republic of Latvia and persons who, for some reason, should have been but were not listed in the Register. The census also recorded inhabitants of

Latvia who had not changed passports for new ones at the critical moment and therefore had no stamp from the Department of Citizenship and Migration Affairs containing an identity code, children not yet registered in the Residents' Register, and persons who had died after the critical census moment.

Pursuant to the provisions of the Law on Population Census, several categories were not included in the census. They were: persons with no identity code and without the registration address of residence in Latvia; persons who were registered in the Residents' Register, but had resided abroad for more than a year; persons who had arrived in the country with an objective to stay in Latvia less than a year; children born after the critical census moment; and persons who had died before the critical moment.

The 2000 Population Census in Lativa was conducted by more than 5,000 people, including 38 assistant managers of regional statistical offices, 430 enumeration supervisors, and 4260 field enumerators.

A part of the information pursuant to the census program—answers to 10 of 33 questions—was obtained from the Residents' Register and from the information system of State Revenue Service.

During the census, 14 questions about each person residing in the dwelling and 9 questions characterizing the dwelling were asked. Questionnaires were completed for each person 7 years of age and older.

In cases where an identity code was used, information on a person that existed in different registers could be combined. Using the identity code also increased the accuracy of the census data and decreased the number of questions asked, thereby lowering the census costs.

From May 11 to June 10, 2000, an important activity to ensure the quality of the population census was conducted. In selected territories, people were repeatedly visited at their dwellings in order to check the correctness of the data in the census questionnaires. The requirement to develop a quality sample survey for the census data was recommended by the UN ECE, "Eurostat," and the Statistical Office of the European Communities. Through mathematical selection, 1 percent, or 42 enumeration areas, were selected to be verified in this manner.

In order to accelerate the processing of census data, the traditional methods—those used for previous population censuses—were not used. Instead, an up-to-date data entry and processing system and software based on optical character recognition were developed with the support of the Latvian and Swedish governments. In order to ensure the operation of this system at the desired speed and with a minimum level of errors, the enumerators were required to write letters and figures in both the household and persons' questionnaires according to a proscribed standard. Using high-capacity, fast-acting scanners, information from 820,000 household questionnaires and more than 2.1 million persons' questionnaires was entered in the computing system. At the same time, using the most recent developments in information technology and the corresponding software, an electronic archives for the images of the basic census documents was developed.

Everywhere in the world, conducting a population census is a highly necessary, but expensive activity. The 2000 Population Census in Latvia cost 2,578,000 lats (4.3 million USD). The per capita cost of the census was 1.08 lats (1.8 USD).

The data processing system developed for the 2000 Population Census has been used to process data in other statistical surveys. It will be used also for short-term entry and processing of other large-scale data arrays.

One year after the population census began on March 31, 2000, the CSB of Latvia published a compilation of the provisional results of the 2000 Population Census. This publication contained data on the population numbers in each administrative territory of Latvia and included 11 tables comprised of data on each of the Republic's seven biggest cities and 26 districts. This involved data on population age and ethnic composition, level of education, native language, skill levels of the most widespread languages, and the main sources of livelihood. Data characterizing the conditions of population housing were published, too. This included housing type, kind of housing ownership, construction period, and facility level, etc.

Information from population census forms was entered into the computer through use of fast-acting scanning and data processing technology. This enabled the first statistical bulletin containing provisional results of 2000 Census to be prepared and printed in a comparatively short time.

After the population census data were processed and verified, statistical tables with the final

summary population data were constructed. These tables, created for each administrative territory in compliance with the Law on Population Census, were sent by the CSB to each municipality free of charge in the first quarter of 2002.

In March 2002, the CSB completed another, more important, phase in the processing of the 2000 Population Census results. On April 5, 2002, "Results of the 2000 Population and Housing Census in Latvia," the main publication containing the aggregated results of the population census was presented to the public, all levels of government and scientific research institutions. This publication has aggregated data on population numbers, sex and age composition, ethnicities, language skills and use, education, main sources of livelihood, families, and number of children born. It also has data characterizing the dwellings of the population.

The majority of indicators are broken down by city and town, district, county, and pagasts. The most important data, furnished in graphs and charts, are compared with the data of the previous census.

This main 2000 Census publication, funded by State budget resources, was sent to all public libraries. It was passed to all subscribers and made available for customers and other users.

Data from the 2000 Census were also placed on CD-ROM, which provides for a wide-range of possibilities for users. In the CD-ROM version, the tables are provided in different formats, such as the PC AXIS software application.

According to the 2000 Population and Housing Census, Latvia had a total population of 2,377,383 persons on March 31, 2000. Since 1989, the total population has decreased by 289,000, or 10.8 percent. The percentage of males in the total population has fallen from 46.5 percent in 1989 to 46.1 percent in 2000.

An unfavorable demographic shows the rapid aging of the population. This is noted by the decreasing number of children and the increasing number and proportion of elderly people. In the period between the two last population censuses, the total number of those 0-14 years of age decreased by 140,000, or 25 percent, and those 15-59 years of age by 184,000, or 11 percent. At the same time, the number of population aged 60 and older increased by 35,000, or 8 percent.

Census data show that the population of Latvia consists of more than 150 ethnicities. According to the same data, one of the most widespread languages of Europe—English, German, or French—is known by nearly every fourth inhabitant of Latvia who has indicated any skill level of a foreign language. The numbers indicate that 340,000 Latvians know English, 179,000 know German, and almost 10,000 know French.

Of the total population of the country, 95.7 percent are citizens and non-citizens of the Republic of Latvia (74.5 percent and 21.2 percent respectively), and only an insignificant number are foreign citizens. Most of the foreign population are citizens of European countries.

The 2000 Census also provided data regarding the educational level of the population aged 15 years and older. Of the total number who responded to the educational level question, the results indicated: 13.9 percent have a higher education, 20.2 percent are secondary specialized, 31.0 percent have a secondary education, 26.5 percent have a basic education, and 5.9 percent have a primary education. Of the same age group, 1.5 percent have a fourth-grade or lower education, 0.6 percent have no official education, and 0.25 percent are illiterate.

The 2000 Census data show that the percentage of married males aged 15 years and older is 57.5 percent, compared with 66.5 percent in the 1989 Census. Married females comprised 51.3 percent of the female populations, as compared with 55.2 percent in the 1989 Census.

The census data show that 68.5 percent of the population reside in dwellings that are the property of a household or other private owner. This is an increase from 19 percent in this category in 1989. The proportion of those living in dwellings belonging to local governments and organizations, however, decreased from 76.9 percent in 1989 to 29 percent in 2000.

Census 2000 provided for a more detailed survey of the amenity level in dwellings. Respondents were asked if they had a kitchen in their dwelling and also if they did not have a shower or bath tub, was a bathhouse available to them. The results showed that of all who indicated the existence of amenities in their dwelling, 98.9 percent have a kitchen, 98.7 percent have electricity, 87.9 percent use a gascooker, 83.2 percent have a waterpipe, 77.0 percent have sewage, 77.4 percent have a flush toilet, 67.9 percent have a bath or a shower, and 56 percent have hot water.

Work is still continuing on the 2000 Census project. By June 30, 2003, 40 summary data tables will be made available to Eurostat. The tables will contain demographic, economic, and social characteristics of the population, as well as provide data on households, families, and dwellings.

The statisticians of the three Baltic countries recently agreed to publish a common data collection, which will have information obtained in population censuses in Estonia, Lithuania, and Latvia. It is envisioned that by the end of 2003 or the beginning of 2004, information users will receive this edition both on paper and in CD format. Moreover, the CDs will contain more data.

While population censuses in Estonia and Latvia were conducted in April 2000, Lithuania conducted its in April 2001. When developing population census programs and lists of questions asked to the population, priorities set by information users in each country, as well as the requirements and methodological recommendations of the UN EEC and European Union were taken into account.

The three Baltic countries are often perceived as a united region in the world. Hence, it is important that data on this region as a whole are available. These publications are perceived to be of great interest both to domestic users and to scientists of other countries and international organizations.

Without doubt, the 2000 Population Census data will serve as a reliable basis for adopting resolutions and developing and implementing programs and plans that are in the interests of every Latvian.

The 2000 Census data certainly will be used as a stable basis and sample framework for other statistical surveys, such as the Household Budget Survey, the Labor Force Survey, and a number of other surveys in the areas of the social and demographic statistics.

Prepared by the Demographic Statistics Division of the Central Statistical Bureau of Latvia, Uldis USACKIS, phone 371 7366900, e-mail uusackis@csb.lv

2001 Population and Housing Censuses in Portugal

Population and Housing Censuses for 2001 in Portugal were conducted on 12 March 2001. The Overall Program for the 2001 Census was approved by the Higher Statistical Council (HSC) and all phases of the work have been followed by this Council, which represents the main statistical users.

The main objectives for these censuses were to satisfy the users' needs by changing the census content and releasing the final data as soon as possible. Special concern also was put on producing and disseminating consistent quality indicators, which should allow users to know and fully accept the coverage and content rates for these censuses.

The Population and Housing censuses of 2001 have followed procedures similar to the censuses of 1981 and 1991 concerning exhaustive collection method for every statistical unit and separate phases for distribution and collection of questionnaires.

Results of the 2001 Censuses were planned to be composed of three types of data, each of which was replaced by the last set released:

- Preliminary results, released on 22 June 2001;
- Provisional results, released on 17 January 2002; and
- Final results to be released by the end of 2002.

The use of scanning and other methods, related to computer assisted coding and automatic imputation, played a central role in achieving those aims.

The questionnaires were scanned by four high level scanners; this work was completed in four months with a scanning rate of 120 2-sided A4 questionnaires per minute by each scanner; the optimized productivity reference for each scanner was 160 per minute.

Results of scanning data capture:

- Numeric characters have recognition rates of:
  - 94.1 percent—immediately recognized; and
  - 5.4 percent—easily recognized in matrix.

- Only 0.5 percent had to be identified/corrected after checking the respective field and other answers of the respective questionnaire;
- Alphabetic characters were processed within the respective alphabetic fields, using dictionaries and an algorithm to rebuild words and alphabetic expressions according to requirements of the respective coding instructions; thus, the rate for alphabetic expressions fully recognized was 76.4 percent with the following distribution:
  - Municipality 84.4 percent;
  - Country 70.4 percent;
  - University field of study 72.2 percent;
  - Occupation 74.9 percent; and
  - Branch of economic activity 68.7 percent.

The remaining alphabetic expressions were recognized with the help of technical staff who were prepared for this task.

Quality evaluation was made mainly through the post enumeration survey. The earliest data from this survey, released in October 2001, points out the net rates of coverage:

- 99.7 percent for buildings;
- 99.5 percent for housing units;
- 100.9 percent for private households; and
- 100.6 percent for resident persons.

The U.S. Economic Census, That "Other" Census

Howard Hogan
Economic Statistical Methods and Programming Division, U.S. Census Bureau

The Decennial Census of Population and Housing of the United States gets most of the attention. However, the U.S. Census Bureau will soon send questionnaires to more than 5 million businesses, launching the 2002 Economic Census. The Economic Census provides the foundation for reports on hundreds of industries and data for states, counties and places. Data covering calendar year 2002 will be collected and processed during 2003, and the first data will be released beginning in early 2004. New questions and new technologies mark advances in a long history of Economic Census improvement.

Congress first ordered census takers to "take an account of the several manufactures within their several districts, territories and divisions" as part of the Population Census in 1810. In 1902, Congress authorized the establishment of a permanent Census Bureau, and at the same time directed that a census of manufacturers be taken every five years. The 1905 manufacturing census marked the first time a census of any kind was taken separately from the regular every 10 years population census. The first census of business, covering retail and wholesale trade, was conducted in 1930, and shortly thereafter was broadened to include some service trades.

The Economic Census was almost cancelled when the Eisenhower Administration failed to provide funding for the 1953 Economic Census. A series of unreserved testimonials from business, financial, professional, and governmental groups led to the reinstatement of the Economic Census for 1954. The 1954 Economic Census provided comparable data across economic sectors, using consistent time periods, concepts, definitions, classification and reporting units. For the first time, an electronic computer (UNIVAC 1) was used to process Economic Census data. The 1954 Economic Census was also the first to be taken by mail, using lists of firms provided from the administrative records of other federal agencies. The automation of the Economic Census dates back to the use of tabulating typewriters in 1900, punchcard tabulating equipment in 1920, and electronic computers in 1954. Since 1963, administrative records have been used to provide basic statistics for very small firms, reducing or eliminating the need to send them census questionnaires Starting in 1967, selected large firms were allowed to file their reports on computer tape.

The 1997 Economic Census was the first major statistical report based on the North American Industry Classification System (NAICS). Developed cooperatively by the U.S., Canada, and Mexico, NAICS replaced the Standard Industrial Classification (SIC) system to provide greater comparability within North America.

New Questions for 2002

Electronic Commerce: The Census Bureau is gathering the first comprehensive information on the e-commerce sales of practically every industry in the United States. E-commerce includes sales, receipts, and/or revenue from any transaction

completed over an Internet, Extranet, Electronic Data Interchange (EDI) network, electronic mail, or other online system. Until these data arrive, we have e-commerce measures for only a few sectors and for the nation as a whole.

Leased Employment: For the first time, the Census will collect information on the use of leased employees at all business establishments, providing new information on a growing phenomenon. Leased employees are those whose payroll is filed with the IRS by an employee leasing company, not by the company where the employee works.

Supply Chain: The 2002 Economic Census will yield data on supply-chain functions performed by manufacturers of goods, those who store and distribute goods, those who transport goods and those who sell and bill for goods. Questions also will identify the extent that certain functions are outsourced to other companies.

New Technology

The Census Bureau is introducing new systems to make it much easier for businesses to report and to get answers to their questions.

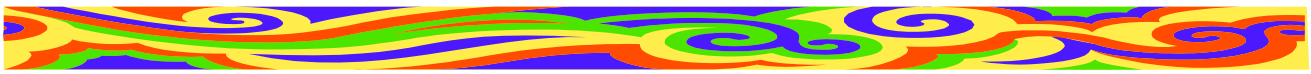
Improved Questionnaires: The 2002 Economic Census will use over 650 different versions of the census questionnaire, each tailored to the characteristics of a specific industry or group of industries. The Census Bureau worked with thousands of trade associations to develop forms that are relevant to business information needs and respectful of business record keeping. Forms have been redesigned so that they are easier to complete and handle. The design and layout of the questionnaires has been accomplished by a sophisticated system based on an electronic “Economic Metadata Repository” and computerized layout system.

Electronic Reporting: For the first time, most businesses will be able to respond by computer. Businesses will still get forms in the mail in December, but, using identifiers on the form, they will be able to download either a questionnaire instrument or a spreadsheet template, complete it on their own computer—importing data from company records directly into the questionnaire—and file the form on a secure Internet site. We consider that allowing businesses to extract data directly from their own spreadsheets is a major breakthrough. Electronic reporting is not new to the Census Bureau. As early as 1967, the agency accommodated a few large retailers who requested permission to file census reports for their many locations on computer tape. In subsequent censuses, the number of requests increased gradually, but the program remained experimental. The 1997 Economic Census made electronic reporting available to thousands of retailers and service sector businesses that requested it. What is different for 2002 is that every business now has the option to report electronically, and the process is much easier and friendlier than ever before. The real payoff is for firms with dozens or even thousands of locations. Instead of having to complete a separate form for each establishment, the business will be able to import information from company spreadsheets right into Economic Census software.

Electronic Help Site: Businesses can learn about their electronic reporting options and get other information by checking the Census Bureau's Business Help Site www.census.gov/econhelp. This new 2002 Economic Census Business Help Site will provide answers to questions businesses ask most often, and other services, such as verification that their report has been received. Companies will also be able to request additional or different forms to be mailed to them. Time extensions can also be requested on-line. Complementing the on-line Help Site is a toll-free "Help Line" to be staffed by Census Bureau employees, twelve hours a day, Monday through Friday.

Data Processing Technology: New optical imaging and data capture technologies will make data entry more efficient and accurate, give census analysts more opportunity to review what is collected, and speed the tabulation of billions of facts about American business.

Publication: The 2002 Economic Census is the second to be published entirely on the Internet and CD-ROM. Only very few reports are published in print, yet portable document format technology allows anyone to print out reports from their computers on demand. Reports and statistics are published on the Census Bureau’s Internet site <www.census.gov>.



Welcome New Members!

We are very pleased to welcome the following new members.

Australia

DAVID STEEL

NICOLE WATSON

Belgium

KADIOBA

Benin

ALEXANDRE BIAOU

Burkina Faso

BOUBACAR TRAORE

Cameroon

ANACLET DESIRE DZOSSA

Central African Republic

ASSANE ABDALLA-KADRE

Chad

DEMSOU THEMOI

Fiji

M. G. M. KHAN

India

NATIONAL SAMPLE SURVEY ORGANISATION

Italy

ANNA GIGLI

FILOMENA MAGGINO

Korean Republic

JUNWOO JEON

MIN KYUNG KIM

Poland

WOJCIECH GAMROT

ALZBIETA GETKA-WILCZYNSKA

Slovakia

EMILIA DURCOVA

Spain

SIERRA JOSU

Sweden

ANDERS HOLMBERG

United Kingdom

JEREMY BARTON

JOHN CHARLTON

JACQUI JONES

SALAH MERAD

United States of America

JENNIFER CZUPRYNSKI

R. SCOTT EVANS

DAVID WHITFORD

DENISE WICKWAR

Zimbabwe

ERICA KEOGH

Announcements

IASS Short Courses for the 54th ISI Session, Berlin 2003

The International Association of Survey Statisticians will sponsor a program of short courses to take place just before and immediately following the 54th Session of the International Statistical Institute (ISI), which will be held in Berlin, Germany, August 13-20, 2003. The courses will be led by international high-level experts in their fields and will offer practitioners, researchers, and students in survey methodology, statistics and social and behavioral sciences the opportunity to improve their skills and discuss issues of common interest. All courses will be presented in English, and participants should have the ability to work in this language. There are some opportunities to sponsor travel and accommodation grants for participants from developing and transition countries. A reasonably priced course fee will be determined and it will cover the costs of all course materials.

The following courses are scheduled to take place just prior to or just after the ISI meetings—during the period August 11, 2003, through August 22, 2003.

Course A: Workshop on Survey Sampling

August 11, 12, and 13, 2003 (2 1/2 days)
Presented by: Graham Kalton and Colm O'Muircheartaigh

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. Topics to be covered: sampling frames and frame problems; various sample designs—simple random sampling, systematic sampling, stratification, cluster and multi-stage sampling, probability to proportional to size sampling; weighting for unequal selection probabilities, nonresponse adjustments, and poststratification adjustments; and an introduction to variance estimation. Those who want to know more on variance estimation are expected to continue in Course B.

Course B: Variance Estimation in Complex Surveys

August 11 and 12, 2003 (2 days)
Presented by: Wayne Fuller and Kirk Wolter

The purpose of this course is to provide training in variance estimation in complex surveys. It covers 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.

Course C: Small Area Estimation

August 11 and 12, 2003 (1 1/5 days)
Presented by: Jon N. K. Rao

Sample survey data can be used to derive reliable estimates for large data areas, but sample sizes in small areas or domains are seldom large enough for direct estimates to provide adequate precision for such areas. This course will provide an overview of indirect estimation methods, traditional as well as model-based.

Course D: Editing and Imputation of Survey Data

August 21 and 22, 2003 (1 1/2 days)
Presented by: John Kovar and Eric Rancourt

Surveys and censuses conducted by statistical agencies, research institutes, and other survey organizations incur various degrees of nonresponse and other data problems even under ideal conditions. To alleviate these problems, editing and imputation methods are usually applied. The aim of the course is to introduce the students to methods of error localization and correction, and prevention, detection, and treatment of nonresponse as well as the evaluation of such methods. Various examples will be given to illustrate the material presented.

Course E: Business Survey Methods

August 11 and 12, 2003 (2 days)
Presented by: David Binder and Mike Hidiroglou

Business surveys are routinely conducted by statistics agencies. They are typically characterized by highly skewed universes that change quite rapidly over time. The challenge is to design surveys that will stand the test of time, and yet produce reliable data. The course presents key issues in statistical and measurement design of business. The course will describe the techniques for designing business surveys, introducing the participant to ad hoc and periodic (ongoing) business surveys. It will cover list frame construction and maintenance, as well as sampling procedures. The instructors will review issues in building and maintaining a list frame (Business Register); sampling procedures, including sample size determination and allocation, and sampling methods; data capture, preliminary editing, and statistical editing; outlier detection and treatment; imputing for missing total and partial nonresponse; and weighting and estimation.

Course F: Designing the Optimal Questionnaire

August 11 and 12, 2003 (2 days)
Presented by: Edith de Leeuw and Don A. Dillman

The course will cover two parts, the first part presents techniques for designing the optimal questionnaire, the second part zooms in on designing modern mail and internet surveys. The workshop will start with a review of basic principles of questionnaire design and expand to optimizing questionnaires for different modes and tailoring for special populations. The focus is on reducing nonresponse and measurement error. The purpose of this course is to provide participants with an efficient and up-to-date tool bag to design and implement quality surveys.

Survey and Statistical Computing IV - The Impact of Technology on the Survey Process

The Association for Survey Computing will be hosting its fourth *International Conference on Survey and Statistical Computing* at Warwick University in the United Kingdom, between September 17 and 19, 2003.

The central theme will be *the Impact of Technology on the Survey Process.* The conference will, *inter alia*, explore the often-complex relationship between the push and pull of technological changes and the expectations and demands created by them. It will also examine how outcomes feed back to affect the processes which initially gave rise to them. Although a conference about technology, the topics to be discussed will appeal to everyone with an interest in survey design, data collection, analysis, reporting, or statistical computing.

Invited speakers will include Norman Glass, (Chief Executive, National Centre for Social Research) and Denise Lievesley, (Director, UNESCO Institute for Statistics). Contributed papers will address such topics areas as: the role of the survey professional, sampling and respondent recruitment, data capture, information dissemination and discovery, software support for the survey process, quality and quality assurance, ethical issues and qualitative research

Complete details about the conference are available from the ASC's website: http://www.asc.org.uk or from
Diana Elder
ASC, PO Box 60, Chesham, Bucks HP5 3QH, UK
Phone & Fax: +44 (0) 1494 793033,
E-Mail: Admin@asc.org.uk

Federal Committee on Statistical Methodology

The Federal Committee on Statistical Methodology (FCSM) is planning a research conference to be held November 17-19, 2003 at the Sheraton Crystal City Hotel, Arlington, VA. The conference will feature mostly contributed papers with formal discussion and software demonstrations. Topics will cover a wide range of subjects including methodology, empirical studies, and issues of concern to the federal statistical community. More detailed information can be found on the FCSM website: www.fcsm.gov.

IASS Program Committee for the Sydney ISI Session (2005) Invites Suggestions

The International Association of Survey Statisticians (IASS) has formed a Program Committee that will shape the program for invited

paper sessions and other activities to be organized by the association during the 2005 period of sessions of the International Statistical Institute (ISI) to take place in Sydney, Australia.

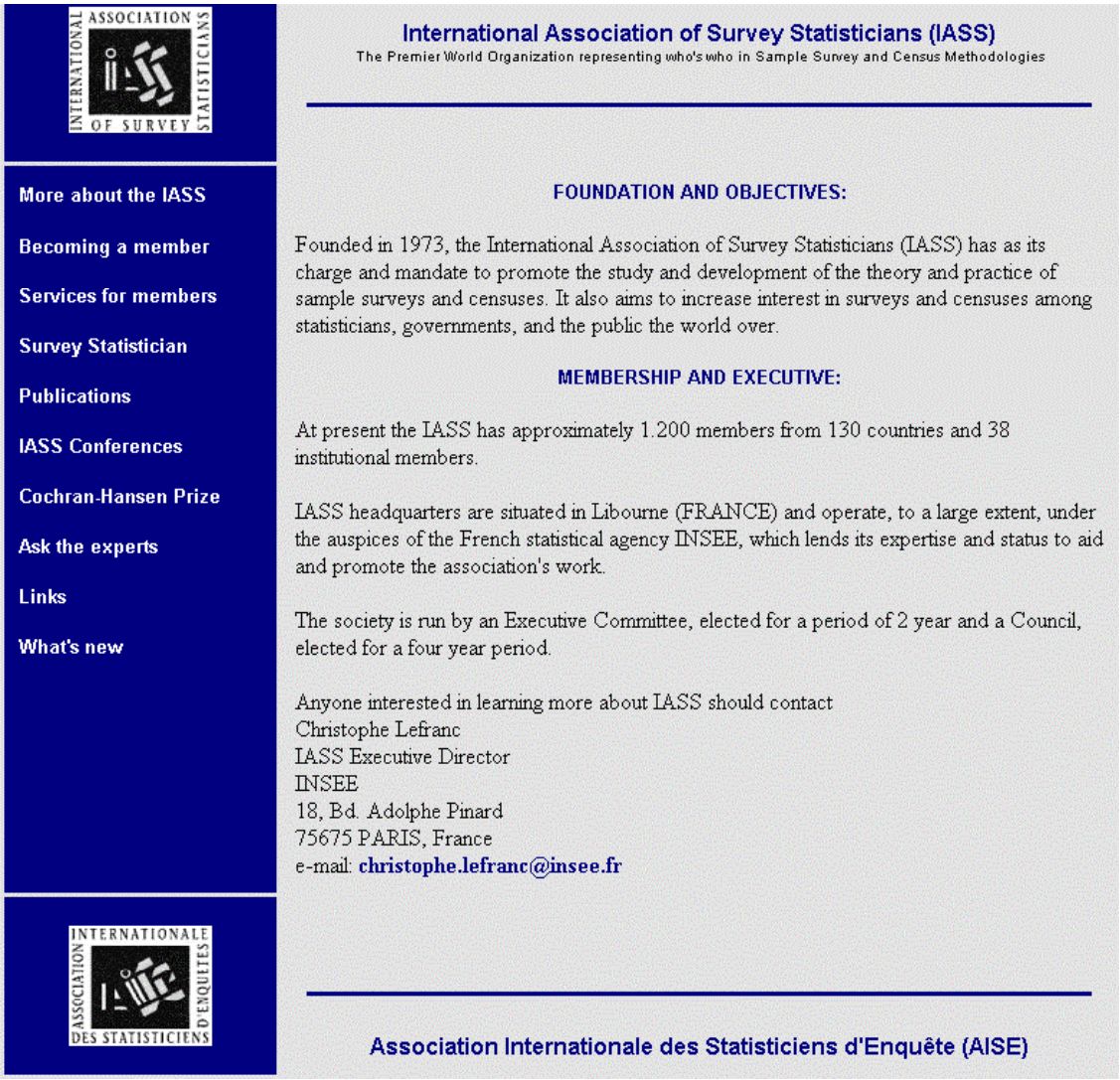
You are invited to offer suggestions of possible topics for invited paper sessions, either to be organized solely by IASS, or to be jointly organized with other ISI affiliated organizations. Proposals should be submitted to the chair of the committee at the address given below before April 30, 2003. Suggestions should, whenever possible, carry also indications of potential organizers for the session.

Pedro Silva

IBGE – Methodology Department
Av. Chile 500 / 10o. andar
20031-170
Rio de Janeiro – Brazil
Phone: 55-21-25144548
Fax: 55-21-25140039
E-mail: pedrosilva@ibge.gov.br

Visit the new and improved IASS web site and read *The Survey Statistician* on line!

www.isi-iass.org



Important Notices

- A PDF file of the newsletter is available on the IASS web site. Very few members no longer wish to receive the hard copy, but simply to be notified when a new issue is posted. Currently, we do not have a process in place to support this option. A process will be developed when an adequate number of members choose the above. Until that time, all members will continue to receive hard copies of the newsletter. Please send an e-mail to LeylaMohadjer@Westat.com if you would like to take advantage of the option mentioned above.
- Members are encouraged to view the IASS website (www.isi-iass.org) and provide comments or suggestions to Fred_Vogel@nass.usda.gov.

In Other Journals

Survey Methodology

A Journal Published by Statistics Canada

Contents Volume 28, Number 1, June 2002

Journal of Official Statistics

An International Review Published by Statistics Sweden

Contents Volume 18, Number 2, 2002

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.

Journal of Official Statistics

An International Review Published by Statistics Sweden

Contents Volume 18, Number 2, 2002 (cont.)

Contents Volume 18, Number 3, 2002

All inquires about submissions and subscriptions should be directed to jos@scb.se

Statistics in Transition

Journal of the Polish Statistical Association

Contents Volume 5, Number 4, June 2002

The Allgemeines Statistisches Archiv (AStA)

www.uni-koeln.de/wiso-fak/wisostatsem/asta/

The Allgemeines Statistisches Archiv (AStA), Journal of the German Statistical Society is edited by Prof. Dr. Karl Mosler, University of Cologne. It provides an international forum for researchers and users from all branches of statistics. The first part (Articles) contains contributions to statistical theory, methods, and applications. A focus is on statistical problems that arise in the analysis of economic and social phenomena. The second part (News & Reports) consists of news, reports, and other material that relates to the activities of the Society. Book reviews are published in the third part (Books). Invited papers which have been presented to the Annual Congress of the Society are regularly published in the second issue of a volume. All papers in the first part are anonymously refereed. In order to be acceptable, a paper must either present a novel methodological approach or a result, obtained by a substantial use of statistical method, which has a significant scientific or societal impact. Occasional review papers are also encouraged. Authors of contributions are requested to mail three copies to the editor for refereeing. The journal publishes original contributions in English and German language, articles preferably in English. After acceptance the author has to provide a LaTeX- or ASCII file of the final manuscript that follows the style of the Allgemeines Statistisches Archiv in citation and layout. For submissions, please contact:

Petra Sablotny
Allgemeines Statistisches Archiv (Prof. Mosler)
Universität zu Köln
Albertus-Magnus-Platz
D-50923 Köln
e-mail: sablotny@marge.statistik.uni-koeln.de

Contents Volume 86, Issue 3, 2002

Articles

The Allgemeines Statistisches Archiv (AStA)

Contents Volume 86, Issue 3, 2002, (cont.)

News and Reports

Books

Contents Volume 86, Issue 4, 2002

Articles

News and Reports

Books

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

Last 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: _____

Fax number: _____

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

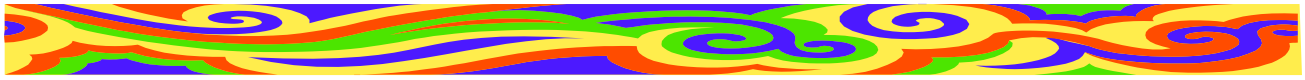
Home ☐ Business ☐

IASS Officers and Council Members

President (2001-2003):	Xavier Charoy (France), xavier.charoy@libertysurf.fr
President Elect (2003-2005):	Luigi Biggeri (Italy), pres@istat.it
Vice Presidents (2001-2003):	David Binder (Canada), binddav@statcan.ca
	Anders Christianson (Sweden): anders.christianson@telia.com
Scientific Secretary (2001-2003):	Seppo Laaksonen (Finland): seppo.laaksonen@stat.fi
Council Members (1999-2003):	Florentina Alvarez (Spain), falvarez@ine.es
	Cynthia Clark (USA), cclark@census.gov
	Pedro Do Nascimento Silva (Brazil), pedrosilva@ibge.gov.br
	Jean Martin (UK), jean.martin@ons.gov.uk
	Vitalis Muba (Tanzania), eastc@ud.co.tz
	Bhagwan Tikkiwal (India), bdtikkiwal@rediffmail.com
Council Members (2001-2005):	Kari Djerf (Finland), kari.djerf@stat.fi
	David Fitch (USA-Guatemala), dfitch@uvg.edu.gt
	Lidija Gligorova (Croatia), lgligor@dzs.hr
	Langhui Huang (China), huanglh@stats.gov.cn
	Moshe Sicron (Israel), sicron@internet-zahav.net
	Awa Thiongane (Senegal-Ethiopia), athiongane@uneca.org

Committees Chairs:	
Berlin 2003 Program Committee:	Danny Pfeffermann (Israel), msdanny@mscc.huji.ac.il
Sydney 2005 Program Committee:	Pedro Do Nascimento Silva (Brazil), pedrosilva@ibge.gov.br
2003 Election Nominations Committee:	Oladejo Ajayi (Nigeria), ooajayi@infoweb.abs.net
Jury of 2003 Cochran-Hansen Prize:	Chris Skinner (UK), c.j.skinner@soton.ac.uk

The Secretariat:	
Executive Director:	Alain Charraud (France), alain.charraud@insee.fr
Treasurer:	François Fabre (France), francois.fabre@insee.fr
Executive Secretary:	Anna Maria Vespa-Leyder (France), vespa@cepii.fr
Secretary:	Claude Olivier (France), claude.olivier@insee.fr



Institutional Members

4 international organizations

AFRISTAT
CICRED
EUROSTAT
UNITED NATIONS STATISTICAL DIVISION

31 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 STATISTIK
FINLAND - STATISTICS FINLAND
FRANCE - INSTITUT NATIONAL DE STATISTIQUE ET D'ÉTUDES ÉCONOMIQUES - INSEE
GAMBIA - CENTRAL STATISTICS DEPARTMENT
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
SPAIN - INSTITUTO NACIONAL DE ESTADISTICA
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