International Statistical Institute

THE SURVEY STATISTICIAN

L'Institut International De Statistique

No. 35
December 1996

Table of Contents

Letter From the President

Dear IASS Member:

Preparing for future conferences has been a major focus of our activity since the last issue of *The Survey Statistician*.

The Istanbul ISI Session which runs August 18-27, 1997 will contain many meetings of interest to IASS members. It is also a great opportunity to meet colleagues from other countries and discuss issues of common interest. If you have not been to an ISI Session before, I would certainly encourage you to attend this one, particularly IASS members from Europe where travel costs may be more manageable. Anders Christianson is the IASS contact for this meeting.

Immediately prior to the ISI Session, we will be holding three workshops

- Variance estimation
- Imputation for missing data
- Quality management for surveys.

More details of these workshops are available elsewhere in this newsletter.

Immediately following the ISI Session will be a joint IASS/IAOS Conference on longitudinal statistics. It will be held in Jerusalem from 27-31 August 1997. Planning under Gad Nathan's leadership is well advanced. It is intended to cover both the design and analysis of longitudinal studies. The focus will be on the special problems posed by the longitudinal nature of the survey. More detail can be obtained from Gad on gad@olive.mscc.huji.ac.il. Additional information is available on the Internet Web Site http://pluto.mscc.huji.ac.il/~gad/smls.html.

Looking at the longer term, Susan Linacre has formed her Committee for the 1999 IASS program. The members are Susan Linacre (Chair), Vijay Verma (India), Sid David (Philippines), David Morgenstein (USA), Danny Pffermann (Israel), Seppo Laaksonen (Finland), Fritz Scheuren (USA) and Anders Christianson (Sweden). I am sure Susan would welcome any suggestions you have on the 1999 IASS program. She has an announcement in this issue that contains more information on the activities of her committee.

A joint IASS/IAOS Conference is being organized for Aguascalientes, Mexico in August 1998. The theme will be "statistics for economic and social development. Instituto Nacional de Estadstica Geographa e Informatica (INEGI) have kindly offered to host the Conference. I recently visited INEGI to make some preliminary arrangements and they proved to be very fine hosts. The INEGI facilities are first class, and a range of accommodation will be available to suit all tastes. A program committee is being established. Mario Palma is looking after local arrangements. If the hospitality they showed me is an indication of things to come, we can expect a fruitful and enjoyable conference.

As you will be aware the 1999 ISI Session will be held in Helsinki, Finland. The ISI would be interested in hosting a satellite meeting either before or after the ISI Session but preferably in the region. If there are any expressions of interest, please contact me.

Talking about ISI Sessions, we will soon be dispatching, to all members, a copy of the IASS invited papers presented to the Beijing Session in 1995.

We are getting closer to having a World Wide Web Site open. We will provide details as soon as possible. It is not accessible at the time of writing but may be by the time you read this letter. The URL for the home page will be http://www.cbs.nl/isi/iass.html.

The ISI Permanent Office will be managing the home page for us to save us the expense and effort of establishing our own operation. Fred Vogel of the National Agriculture Statistical Service has agreed to be the editor for the home page and any material should be submitted through him. Please do so and help make the IASS home page a success. The preferred method is by diskette in ASCII format although E-mail would also be acceptable.

It is also possible for our members to access SRMSNET under a special arrangement organized by David Binder who is the current chair of the Survey Methods Section (SMS) of the American

Statistical Association. SRMSNET is a free service. To subscribe to SRMSNET, simply send a message to

listserve@umdd.umd.edu

and in the body of the message, type

subscribe SRMSNET *John Doe*

substituting your own name for *John Doe*. Within 5-10 minutes you should receive confirmation of your subscription and instructions on how to post or get information.

To conclude this note, I would like to again say something about the future of survey methodologists. A very important paper by Don Dillman is soon to be published in the Journal of Official Statistics. He is questioning the effectiveness of methodologists in official statistical agencies because of the large cultural differences between the methodologists and the operational arm of agencies. These differences make it very difficult for the two groups to communicate effectively.

Whilst cultural differences undoubtedly exist, it is a very important part of our job to bridge those differences. We will become irrelevant unless we do so. We belong to a practical scientific discipline — we have to communicate with the practitioners.

How do we improve that communication? I have listed below some personal suggestions.

- Ensure there is support for methodological activities from the very top level of the organization.
- Build the confidence and trust of key operations staff — physically outposting methodology staff to operations areas is one effective way of doing this.
- Get a real understanding of practical issues (and the operations perspective). It is a great learning experience (and will improve their future effectiveness) for methodology staff to work in the operations area for a period of time.
- Share research results with operations staff in a language that can be understood.
- Ensure the priorities of research work are supported by operations staff.
- Build your networks around those operations staff that have some understanding and empathy with methodology issues.

I hope you all have a great festive season and enjoy 1997.

Dennis Trewin
IASS President

.....

A Content Analysis of Advance Letters From Expenditure Surveys of Seven Countries

Martin Luppes

The three main reasons for sending out advance letters are: to announce the visit of the interviewer; to give information about the survey; to stimulate the willingness of the respondent to participate in the survey. In most surveys the information in advance letters covers the objectives of the survey, its benefits and the costs to the respondent.

In this study the contents of advance letters are compared with respect to the persuasive arguments used to stimulate participation in surveys, as regards topics covered as well as compliance principles explicitly referred to. A conceptual framework is used which describes the relation between the content of an advance letter and the way in which a respondent is stimulated to participate. The content of an advance letter can be described by three dimensions:

- Background information such as the objectives of the survey and the name of the responsible survey organization;
- Benefits of the survey, including direct benefits;
- Costs of the survey such as burden of the interview.

Within each dimension several topics are distinguished. In total twelve different topics are used in the analysis (Luppes, 1995).

However, the heuristic decision whether or not to comply with a request is made according to the so-

called compliance principles that individuals use as a basis to decide when it is appropriate and adaptive to comply. In this study seven compliance principles are distinguished:

- RECIPROCATION, the tendency to participate if a reward is given;
- CONSISTENCY, the tendency to behave in a similar way over situations that resemble one another;
- SOCIAL VALIDATION, the tendency to behave according to the norms and values of the social group to which one belongs (or thinks one belongs);
- AUTHORITY, the tendency to be more willing to comply if the request comes from an authority;
- SCARCITY, the tendency to comply because the respondent gets the feeling of being in a unique position.
- LIKING, the tendency to comply with requests from attractive requestors.
- HELPING, the extent in which people in general are willing to help when asked.

To analyze advance letters the content of an advance letter and the compliance principles can be represented in matrix-form, such that the rows represent the various content topics and the columns the compliance principles. Each cell in this matrix then represents a relation between a topic in the letter and a compliance principle. This matrix is the basis of the coding form that was scored for each of the seven letters by three independent researchers (judges) from Statistics Netherlands. The study used the scores given by two of the three judges.

Advance letters used in Family Expenditure Surveys (FES) from seven countries were selected for this study (letters from USA, Belgium, Great Britain, Finland, France, Netherlands and Sweden). To reduce and standardize the amount of material, all enclosures to these advance letters were omitted.

In most letters the choice of topics is limited to the objectives of the survey, the name of the survey organization, the direct benefits for the respondent if he or she participates and the promise of anonymity and confidentiality. The use of compliance principles in the letters is almost totally limited to the principles of reciprocation (often in combination with a remuneration), authority (with reference to the law) and helping tendency (a general call for help). Differences between the letters are mostly due to the fact that these compliance principles are used in different ways. In general, explicit referral to compliance principles appears to occur rather haphazardly and infrequently.

For assessing advance letters the coding-form appears to be a worthwhile instrument. Attention can be drawn to the salient topics in a letter and the way these topics are formulated, *e.g.* the persuasiveness of the message and wording. As a checklist, the coding-form may be of help for the design of advance letters.

The interpretation and evaluation of the letter, however, should be explored by other means, *e.g.* by in-depth interviews in which thinking aloud and paraphrasing are used. The information from these interviews, together with insights from content analyses, may lead to a better and more systematic understanding of the role of advance letters.

For further information contact: Martin Luppes at Statistics Netherlands E-mail: MLPS@cbs.nl.

Reference

Luppes, M.J.G., A Content Analysis of Advance Letters from Expenditure Surveys of Seven Countries in: Journal of Official Statistics, Vol. 11, No. 4, 1995, pp. 461-480.

.....

Some Issues of Sampling Surveys in Poland in the Transition Period

Jan Kordos

The activities on the application of sampling methods in the sense of random selection of a sample were started in Poland at the beginning of 1930's by J. Neyman, who published the results of his work in Polish in 1933 (Neyman, 1933). He

published the main ideas included in this work in English in 1934 (Neyman, 1934) in a famous article devoted to stratified sampling and the method of purposive selection. As is well-known, this work had a crucial influence on the widespread application of the sampling method in the sense of random selection of a sample. At this time, another outstanding Polish statistician J. Piekalkiewicz (Piekalkiewicz, 1934) closely collaborated with J. Neyman and prepared reports from surveys designed by Neyman, analyzed them, and interpreted the results. In 1994 an international scientific conference devoted to the 100th birthday of J. Neyman was held in Warsaw, at which his contribution to the development of statistics, especially his influence on the development of the theory and practice of sample surveys, was discussed.

Despite an early start of activities on the implementation of sampling methods in Poland, the actual use of this method in the Polish statistical practice was at the beginning very limited. Before World War II the application of sampling methods was limited to the acceleration of the generalization of the results of the 1931 Population and Housing Census and some experiments in the field of social surveys (Szulc, 1967).

Sample Surveys in the Centrally Planned Economy

After World War II the Central Statistical Office (CSO) of Poland started activities on the implementation of sampling methods in statistical surveys. In 1949 the Mathematical Commission of the CSO was established (Kordos, 1975) as an advisory and opinion-making body to the CSO President in the field of sample surveys. It concentrated specialists in sampling methods both from the CSO and research centers in the country. Since 1966 the Section for Mathematical Methods of the CSO Statistical and Economic Research Center had a considerable influence on the development of sample surveys. It prepared sample designs, determined the sizes of sample, the estimation methods, and the estimation of sampling errors. It should be added Prof. J. Neyman consulted on the sample surveys during his visits in Poland (Fisz, 1950; Zasepa, 1958). Professor J. Neyman made the Polish statisticians interested not only in the issues related to sample surveys, but also mathematical statistics, the theory of probability, and the general theory of statistics.

However, under the social and economic system in Poland after World War II and prior to 1989, statistical surveys were based mainly on complete statistical reporting, complete housing and population censuses and agricultural censuses. Sampling methods were used in a limited scope. Sampling methods were used mainly in social surveys (household budget surveys and living conditions surveys) (Kordos, 1985), in agricultural surveys (agricultural sample censuses and sample surveys on agriculture (Kordos, 1968), to accelerate generalization of the results of population and housing censuses in the years 1950 and 1960, sampling also was used in 1974 and 1984 in the population and housing censuses, the so-called micro-censuses (Bracha et al., 1979, Zasepa, 1993). Sampling methods were hardly ever used in economic surveys. Despite those limitations, sampling methods were developed in Poland and were taught at higher economic schools, and in 1962 the first in Poland handbook on sampling methods was published (Zasepa, 1962). Later on five more books on sampling methods were published. A few Polish statisticians worked as international experts in sample surveys in various countries hired by specialized agencies of the United Nations.

The remainder of this article is an overview of issues related to the widespread application of sampling methods in statistical surveys and the difficulties encountered, in the period of transition of Polish statistics and its adjustment to the requirements of a market economy which occurred after 1989.

Sample Surveys in the Transition Period

Despite considerable experience in the application of sampling methods in the previous period, there were considerable difficulties in the application of these methods on a larger scale. The biggest difficulties were in economic surveys, which in the previous system were based mainly on complete statistical reporting.

The difficulties are especially acute in ***data for small areas***. Since there was complete statistical reporting under the previous system, small area

estimates were not a major problem. That is why Polish statisticians are so interested in the issues related to the estimates for small areas. We organized an international scientific conference on methods of data estimation for small areas in Warsaw in 1992 (Kalton, Kordos and Platek, 1993). Activities in this field are being continued.

Another important and difficult problem is obtaining adequate ***sampling frames*** in economic surveys. In the transition period, many small-size enterprises came into existence and there were no up-to-date registers which could be used as sampling frames. This universe constantly changes over time, new enterprises emerge, those existing for a short period of time are liquidated. In addition, we observe ***considerable rates of non-response*** which make generalization of the results for the whole universe difficult.

It is much easier to apply sampling methods ***in social surveys*** due to the past experience (Kordos, 1985). The research tradition and previously conducted experiments in the field of ***household budget surveys*** made it possible to make further improvements in these surveys and to adjust to international standards (Kordos, 1995). Surveys ***on poverty and monitoring of the living conditions*** of Polish households were started (Panek, 1996), and for the first time in 1992, a ***labor force survey*** was launched (Szarkowski and Witkowski, 1994) in accordance with the ILO recommendations in this field. These surveys are continuing with success. These surveys were considerably improved by adding new modules periodically which make more in depth analysis possible. Considerable assistance was provided from other countries, especially from the U.S. Bureau of Labor Statistics and the French INSEE.

Although there was some experience in the previous period in the application of sampling methods in ***agricultural surveys***, activities were undertaken to adjust agricultural surveys to international standards in collaboration with the U.S. National Agricultural Statistics Service, (Skow and Wanke, 1994). These activities are being continued. In 1996 a ***complete agricultural census*** was conducted which will be used, inter alia, for the creation of ***sampling frame for agricultural surveys***, more efficient estimation, and the improvement of the system of agricultural surveys.

Recently activities have also started on the improvement of the ***methodology of panel surveys*** which refer to statistical surveys in various fields. These are difficult surveys which require a specific approach. Special training was prepared for selected CSO employees and research work was undertaken.

We believe the staff of CSO should get acquainted with sampling methods, their advantages and disadvantages. An adequate system of statistical training is required which will enable us to provide training for the core statistical staff. This training should be treated as a long-term process in connection with similar training provided within the international cooperation activities, especially the Training of European Statisticians. The training should be consistent with the current international practice of sample surveys, make possible critical analysis of particular stages of surveys, to lead to more effective designs for surveys in the future and for the improvement of already conducted surveys. For these reasons, a new system of statistical training is under preparation in Poland in which issues related to sample surveys are treated extensively.

In addition to systematic training, the next important precondition is conducting research activities related to both theoretical and practical issues. First of all, it would be worthwhile to analyze the usefulness of various sample designs and sampling frames, methods of estimation, weighting and imputation procedures, the relation between the costs of surveys and precision of the results, use of data from other sources, use of unconventional methods of estimation of data for small areas, methods of data collection for reducing errors, application of estimates and data adjustments, ways of analysis of the precision of the results, *etc.*

There is an urgent need to promote sampling methods among a wide circle of statisticians and the users of statistical data and inform them about the basic research techniques, as well as the way of drawing conclusions on the basis of the results of sample surveys. There are a lot of myths and misunderstanding concerning sample surveys which should be clarified.

For more information contact: Jan Kordos at Al Niepodleglosci 208, 00-925 Warsaw, Poland.

References

Bracha, Cz., Kordos, J., Kowalska, H., and Lednicki, B. (1979), Mikrospis ludnoœci 1974 (*The 1974 Micro Census of Population*). In: Metodologia badañ reprezentacyjnych GUS *(Methodology of Sample Surveys in CSO)*, "Biblioteka Wiadomosci Statystycznych", vol. 29, Warszawa.

Fisz, M. (1950). Konsultacje prof. Neymana i wnioski z nich wyplywajace (Consultations with Prof. Neyman and Conclusions). Studia i Prace Statystyczne, nr 3-4.

Kalton, G., Kordos, J. and Platek, R. (1993). Small Area Statistics and Survey Designs, Vol. I: Invited Papers; Vol. II: Contributed Papers and Panel Discussion. Central Statistical Office, Warsaw.

Kordos, J. (1968), Zastosowanie metody reprezentacyjnej w statystyce rolniczej (Application of sampling methods in agricultural statistics), "Przeglad Statystyczny", nr 3.

Kordos, J. (1975). 25 lat dzialalnosci Komisji Matematycznej GUS (25 years of activities of the CSO Mathematical Commission), Przegl'd Statystyczny, z. 1.

Kordos, J. (1985), Towards an Integrated System of Household Surveys in Poland, "Bulletin of the International Statistical Institute", (invited paper), vol. 51, Amsterdam, Book 2 , p. 1.3.1 - 18.

Kordos, J. (1995): Nonresponse Problems in Polish Household Surveys, Statistics in Transition, vol. 2, number 5, pp. 789-812.

Neyman, J. (1933). Zarys teorii i praktyki badania struktury ludnosci metoda reprezentacyjna (An outline of the theory and practice of representative method applied in social research). Instytut Spraw Spolecznych, Warszawa 1933.

Neyman, J. (1934): On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. Journal of the Royal Statistical Society, 1934, p.558-606.

Panek, T. (1996), A monitoring system of the living conditions of the Polish households. "Statistics in Transition", vol. 2, No. 4, p. 979-1003.

Piekalkiewicz, J. (1934), Sprawozdanie z badañ skladu ludnosci robotniczej w Polsce metoda reprezentacyjna na podstawie materialów spisu powszechnego ludnosci w 1931 r. (Report on Investigation of Composition of Workers' Population by Sampling Method). Instytut Spraw Spolecznych, Warszawa.

Skow, D. and Wanke, H. (1994). Testing an area sampling frame for agricultural statistics in Poland. "Statistics in Transition", vol. 1, No. 6. p. 797-810.

Szarkowski, A. and Witkowski, J. (1994), The Polish Labor Force Survey, "Statistics in Transition", vol. 1, No. 4.

Szulc, S. (1967), Metody statystyczne (Statistical Methods), PWE, Warszawa.

Zasepa, R. (1958): Problematyka badañ reprezentacyjnych w GUS w swietle konsultacji z prof. Neymanem (Problems of Sampling Surveys of CSO in the light of consultations with Prof. J. Neyman), Wiadomosci Statystyczne, nr 6, p. 7-12.

Zasepa, R. (1962): Badania statystyczne metoda reprezentacyjna (Statistical Surveys by Sampling Method), PWN, Warszawa.

Zasepa, R. (1993), Use of Sampling Methods in Population Censuses in Poland, "Statistics in Transition", vol. 1, No. 1, p. 69-78.

.....

Computing Upgrade for Statistics New Zealand

Keith Sykes and Mike Moore

Introduction

Statistics New Zealand (SNZ) is currently undertaking a significant upgrade of its computing environment. The main emphasis of the upgrade is a shift from operating a large, nearly obsolete mainframe computer, towards an open system

computer environment using client server technology.

Associated with this technology upgrade is a strategy of replacing existing systems, characterized by their rigid software design, limited functionality and high maintenance overheads, with easy to use, modular and flexible client server applications.

The development of a new Business Register and Classification system are seen as critical projects in the overall Information Technology (IT) strategy and a pre-requisite for the redevelopment of the many associated systems. These two areas are discussed below.

Business Register Redevelopment

The project to redevelop the Register commenced in October 1994 and is scheduled for implementation in December 1996. The new Register has been developed using Gupta's SQL Windows as the on-line application software and SAS for bulk processes. Sybase System 11 is used for the database. SSA Name3, a proprietary searching engine, is used for name and address searching and matching.

The new Register will run on a Novell Netware Wide Area Network distributed over five sites within New Zealand. The client PC workstations use Microsoft Windows 3.11 as an operating system, although an upgrade to Windows 95 is intended soon after deployment of the new Register. The Sybase database is installed on an Intel based server which runs on a Microsoft NT Server operating system.

The new Business Register system will have several design features that distinguish it from the existing mainframe based system.

- **On-line interactive updating**, which will allow updates to be applied to the database at the push of a button. This replaces a system which is transaction based with overnight updating.

- **Drag and Drop Business Structure Maintenance**, which provides a way of graphically transferring selected units between Business Structures in a matter of seconds. This replaces what was a slow two stage process run over two nights.

- **A History Facility** which captures the changes occurring to Register variables, together with the date, the reasons why the change occurred and the source of that change. The history facility can display a list of all changes for a particular Register unit. Alternatively it can display a field by field view of history.

- **Chart Display of Business Structures** which allows any given business or group structure to be displayed in terms of its component statistical units in a top-down hierarchy. This is an entirely new feature which was not possible with older technology. The visual picture allows the user to see an entire structure at a glance. Structures can be created or changed using this feature also.

- **The Linking of Register Units to Tax Data** which allows the user to carry out full name and address searching on tax (and other administrative data in the future) from within the Register System, instead of accessing these data via stand alone systems.

- **Ownership Links** which will allow associates to be recorded as well as subsidiaries to which the existing system is restricted.

- **A Name and Address Searching Facility** which scores the matches and returns the most probable matches to the top of the list. It also casts a wider net than the existing system, as it is able to find units on the register even when the operator misspells the name.

Expected Benefits

Features such as on-line real time updating and drag and drop structure maintenance are expected to result in resource reductions of nearly 25 percent of the staff currently engaged in register maintenance.

The move from a centralized Register updating service to a single point update approach will see

survey operators and respondent managers updating the register directly. This first point of respondent contact approach will provide a significant reduction in the double handling of register update information from respondents. This approach will allow register staff to focus its resources on the large complex units and quality monitoring and measurement activity.

A consistent reference to standard classifications will be ensured with the new Register's very tight integration with the new Classifications and Related Standards System (CARS) and economic and financial surveys.

The Register links to all economic and financial survey data will mean that survey designers will have a much wider range of size information (turnover, capital *etc.*) available to them for designing business surveys.

The direct link to the tax system will allow more effective use of tax data to reduce survey and compliance costs on small businesses and reduce frame maintenance costs. In addition, access to 1.4 million tax records provides SNZ opportunities to produce a wider range of statistics on small businesses.

The Register's History Facility will make it easier to monitor changes and analyze the reasons for those changes. In particular, it opens up opportunities to produce a wider range of demographic and change statistics about NZ businesses. Backcasting survey data will also be made simpler through this facility.

The Ownership linking facility which allows associate ownership details to be recorded as well as the traditional subsidiary is seen as a significant step towards integrating our Balance of Payments Surveys with the Register. This feature, coupled with the Register's ability to maintain filter information on overseas activity, has set the scene for full integration within the next 12-18 months. The completion of this will see SNZ running all its economic and financial surveys off one integrated Register System.

Classification System Development

Whilst many statistical organizations have over the years developed some form of centralized system for holding classifications, this is a significant new development in Statistics New Zealand. Considerable assistance in the planning stages was received from Australia, France, Canada and Finland in the design of SNZ's system. An early version of EMMA (being developed by World Systems (Europe) Ltd) was also evaluated prior to the system development work commencing. Further cooperation continues with the Australian Bureau of Statistics in the supply of a computerized coding system (WinSCM) which comprises part of the project.

The Classifications system is now in production with some 300 plus social, economic and geographic classifications and concordances currently being loaded to the system. This includes some 400 classification versions.

The purpose of CARS is to use database technology to provide centralized classification storage and maintenance facilities for use in the development and processing of surveys and in the subsequent analysis and evaluation of the data.

CARS will contribute to these goals by: improving efficiency in survey operations, increasing reliability in survey outputs, using innovative database technology to meet client needs, and contributing to the long run integrity of the statistical system.

CARS together with other metadata will play an important role in helping SNZ meet one of its objectives of defining and promoting the use of standard statistical concepts, procedures and classifications thereby enhancing the relatability of data from various survey outputs.

The CARS system has been developed using a traditional three-tier approach. The first tier being the database, the second the access procedures and the third the user interface. The system has been developed using Gupta's SQLWindows with Sybase System 10 as the repository. An Australian product, WinSCM, is being integrated for computer assisted coding. CARS is intended for use throughout the department, encompassing five sites nation-wide. The client workstations are currently Windows 3.1, but can be any of the Microsoft family. The Sybase server is an Intel based Windows NT server, with Novell Netware being used as the Network architecture for the TCP/IP Wide Area Network connecting the sites.

Features of CARS

Child-parent category relationships are included in classifications. This means that CARS is able to store hierarchical classifications, even when the hierarchical structure is not evident from the coding system.

An important feature of CARS is standardized classification maintenance facilities to update or revise classifications as necessary. When a classification is updated or revised, the previous version remains stored, and a new version created. Categories in classification versions can be created, updated or deleted on-line. They can also be loaded from external text-formatted files. Concordances can also be updated, but they are not versioned. Concordances can be used to precisely document category changes between versions of a classification, *i.e.* to provide code histories.

Any person in Statistics NZ can browse on-line released classifications and concordances stored in CARS. Finding particular classifications is facilitated by "drilling down" to them through a subject/topic hierarchy. Every classification is assigned a topic, such as industry, which is the particular theme of that classification, and topics are assigned to broad subject matter areas. If the browser already knows the short name of the classification, a short-cut facility brings it straight up on the navigator screen. Once the version of the classification to be looked at has been selected, the various categories and their properties (code, descriptor, *etc.*) can be viewed in hierarchical order or by level.

Survey system applications will be able to directly access CARS through standardized access routines. New and existing surveys and data processing systems, as they are developed or redeveloped, will use CARS.

Concordances can be viewed in forward or reverse order and there is provision for category pair "translation aids" (for example, percentage breakdowns where a source category does not translate to one target category).

Classifications and concordances in CARS are owned by owner user groups, which contain an individual or a group of individuals. An owner user group is responsible for maintaining the classifications and concordances it owns. There is also a CARS supervisor group which makes sure that classifications are assigned an owner user group and are given the right short name and topic, before the owner puts them into CARS.

All classification versions and concordances go through a life cycle, from set up in progress to verified to released to obsolete to (possibly) deletion. The version verification process checks category properties for consistency. A concordance is verified against the underlying versions.

Version and concordance reports can be obtained. Category differences between versions (whether of the same or different classifications) can also be reported. An administrative report on CARS's holdings of classification versions is also available.

Category synonyms can be stored for classification versions. Codefiles are created by a centralized facility for computer assisted coding. Classifications stored in CARS for commercial clients can be kept confidential if desired.

System Benefits

CARS is SNZ's official central repository for all economic, social and geographic classifications, concordances and codefiles, for use by all departmental data collections. Some of the benefits arising from this are described below.

- Common storage facilities will make it easier to locate and access classifications, concordances and codefiles, benefiting the survey development process. Common browsing, retrieving and updating procedures for classifications and concordances in CARS help reduce the time and resource required to operate surveys.

- Common procedures in CARS for maintaining these classifications and concordances eliminate the need to provide separate maintenance facilities in departmental survey systems. When a classification is updated or revised, the previous version remains stored, and a new version created.

- Having all the classifications in one place and in one format facilitates the use of standard classifications in surveys. The lack of a centralized classifications system can result in surveys utilizing their own variations of standard classifications, *i.e.* non-standard classifications. CARS will contain both standard and non-standard classifications. Increased use of standard classifications will result in closer integration of data from various surveys. This will ultimately result in the department being better able to satisfy customer requirements.

- Classification and Standards Section will be better able to expand their service to SNZ in terms of classification advice and timely promulgation of new and revised standard classifications, concordances and codefiles.

- Many SNZ divisions report a need for centrally held but universally accessible concordances. Much of the time-consuming work in producing standard and non-standard outputs has to do with finding whether concordances exist or with manually constructing concordances because one can't be found (even though it may exist somewhere else in SNZ). In addition to concording different classifications, it will precisely document the changes between versions of *one* classification. This will facilitate the provision of time series analyses.

- The replacement of manual coding of surveys by centralized computer assisted coding facilities will considerably reduce the overall resources required to code survey responses, and improve coding accuracy and consistency. A standardized computer assisted coding facility will be available to external users.

- CARS will also result in considerable saving in staff training and familiarization, especially in areas with relatively high staff turnover. Survey processing, research and programming staff will only need to learn to access one classification system rather than many different ones.

- A library of standard labels for output will result in increased standardization and improved presentation of tables in the Department's publications. CARS verifies classifications before they are released. This will reduce the risk of error from inadvertently using corrupt classifications. SNZ's ability to market classification maintenance facilities will be enhanced, especially as particular classifications can be kept confidential to the owner or to specified readers in CARS.

For further information on the Business Register Development, contact: Keith Sykes, Statistics New Zealand, Fax: +64 9 379 0859. For further information on the Classification System Development, contact: Mike Moore, Statistics New Zealand, Fax: + 64 3 374 8701

.....

Software Review

(Software Review is a new section to appear in *The Survey Statistician* featuring articles on software specifically designed for the analysis of survey data. The focus of the series is on programs that properly estimate standard errors for estimates accounting for complex design and estimation schemes. We would like to thank Jim Lepkowski for accepting the job of section editor for the series. Ed.)

.....

Sampling Error Software for Personal Computers
Jim Lepkowski and Judy Bowles

Many students in sampling methods courses are often surprised to learn that cluster sampling generates positive correlations among sample elements and consequent losses in precision relative to the simple random selections they have been trained elsewhere to analyze. Their surprise turns to dismay when they learn that the standard statistical analysis packages that they use routinely do not account for this lost precision correctly.

One of the next questions they ask is, "What software is there that does take the cluster sampling into account in an analysis?" The appropriate answer depends somewhat on how

much material has been covered in the course about design based inference. The methods of inference from complex samples and the statistical techniques that support them are covered in several textbooks and monographs, but those topics are often more than can be covered adequately for students in an introductory sampling methods course. A satisfactory initial answer to the question about software for design-based survey inference may be a brief catalogue of suitable packages.

Our purpose in this article is to give such a catalogue. We are interested in currently available statistical estimation software which accounts for clustering and other complex features of the stratified multistage samples. We do not attempt to give even a limited treatment of the inferential and analytic methods that underlie this software. Readers interested in the inferential basis or particular techniques might examine the monograph by Lee, Forthofer, and Lorimor (1989), or study the more advanced treatments given in Skinner, Holt, and Smith (1989) or Lehtonen and Pahkinen (1995).

The small catalogue of suitable software given here is further limited to those programs or packages that are currently available on personal computers. Thus, we do not discuss one of the earliest packages which provided such software, OSIRIS (Institute for Social Research, 1982), which offered a suite of programs which accounted for complex design features in estimation for means, proportions, and regression statistics, because the OSIRIS sampling error software is only available at present for mainframe computers.

We list only commercial or documented free-ware statistical software packages that are currently available for use by the general survey data analyst. The eight packages catalogued here are CENVAR, CLUSTERS, Epi Info, PC CARP, Stata, SUDAAN, VPLX, and WesVarPC. There are no doubt other packages that are commercially or otherwise available to the general survey analyst that we have failed to list. We apologize for our oversight, and welcome the opportunity to list other software in a future issue of *The Survey Statistician*.

Most of these packages have more extensive features than only estimation for complex sample survey data, incorporating software for processing and managing survey data as well. The presentation does not attempt to address these more extensive features, concentrating on facilities for handling complex design features. These packages are described in alphabetic order following a brief description of specific features. We conclude with a few remarks about sampling error estimation, the more commonly used statistical analysis software packages, and a brief listing of features we think sampling error software ought to incorporate in the future.

Sampling Error Estimation

Unequal probabilities of selection and compensatory weights, nonresponse and noncoverage compensatory weights, population control adjustment, poststratification, stratification of sampling units, and multistage selection are widely used features in "complex sample designs". At a minimum, survey data must have a weight, stratum identifiers, and sampling unit identifiers for each responding sample element. Sampling error estimation software must have the capacity to account for weights, stratification, and sampling unit in the estimation process.

Nearly all standard statistical software will handle weights correctly for estimating point estimates in most analysis provided. Few handle the weights correctly for variance estimates. Further, only one, Stata, has estimation features to account for the stratification and multistage selection employed in the design.

The programs listed in the next section all report handling weights, stratification, and multistage selection correctly in estimation of point and variance estimates. They all require the specification of weights, strata, and sampling units for each sample element. They do not all handle every conceivable sample design in an unbiased fashion. For example, primary sampling units in most stratified multistage sample designs are selected with probabilities proportionate to size and without replacement. Only one program in the list, SUDAAN, has features to handle explicitly this type of design. However, all listed programs will handle such a design under an ultimate cluster sample selection model (Kalton, 1979). Under the ultimate cluster sampling model, elements within primary sampling units are divided into ultimate clusters and a without replacement selection of those ultimate

clusters is chosen across primary sampling units. Variance estimates are computed using only between first stage unit totals without having to compute variance components at each stage of selection. All programs use the ultimate cluster sampling model in variance estimation, although SUDAAN also has features to estimate variances for designs employing without replacement selection of primary sampling units.

There are two principal methods of variance estimation employed for complex sample designs: Taylor series approximation and repeated replication (see Wolter, 1985). Repeated replication methods include the method of random groups, the jackknife, balanced half samples, bootstrap methods, and various modifications of these methods. The Taylor series approximation and the repeated replication methods do not produce identical estimates of sampling error, but empirical investigation (Kish and Frankel, 1970; Kish and Frankel, 1974) have shown that the differences in the methods for many statistics is small. The important practical implication is that a Taylor series approximation must be developed analytically for each statistic, while repeated replication uses the same basic estimation method regardless of the statistic being estimated. Most of the listed programs use Taylor series approximations to compute sampling error estimates. The repeated replication programs in the list offer many of the basic methods, except the bootstrap.

A variety of statistical methods are handled by the listed programs. Some estimate sampling variances and related statistics (design effects, misspecification effects, intracluster homogeneity) only for means, totals, and proportions for the total sample, for subclasses of the total sample, and for differences between subclasses. Others estimate sampling variances for regression and logistic regression statistics. Nearly all estimate test statistics based on these sampling variances. A few compute variance estimates and associated test statistics for survival analysis, contingency table analysis, generalized estimating equation models, specialized ratio estimates, and standardized rates.

Most of the listed programs operate under DOS. Two of the DOS-based programs (CENVAR and Epi Info) use menus to display options for users to specify options. The other DOS programs (CLUSTERS, PC CARP, VPLX) use keyword input. There are also two programs that operate under Microsoft Windows. One of these programs (WesVarPC) employs "pull down menus" for user specification of program options, while another (SUDAAN version 7) employs a combination of pull down menus and keyword input.

We have not attempted for this brief catalogue to compare the estimates produced by these programs, except in limited testing to be sure the software could be loaded and operated successfully on our microcomputers. A small number of cross-tabulations were prepared from a modest sized sample survey data set (n = 3,617) with all programs except VPLX and PC CARP. Estimated proportions, standard errors, and coefficients of variation were similar for all the programs in this testing. More extensive comparisons are needed across sample surveys of different sizes and for more statistics, but those comparisons are beyond the scope of our current review.

Finally, five of these packages (CENVAR, CLUSTERS, Epi Info, VPLX, and WesVarPC) are available for free, or at a nominal charge to handle processing and shipping. Access is made even easier for all but CLUSTERS since they can be downloaded over the World Wide Web (WWW) using standard network browsing software. Information is available over the WWW for all but one of the programs. In the listings below, E-mail addresses to obtain further information are given for all software, and WWW access to information or to the software and documentation itself is also given.

Sampling Error Software

CENVAR (U.S. Bureau of the Census; contact International Programs Center, U.S. Bureau of the Census, Washington, DC 20233-8860; E-mail IMPS@census.gov; WWW access to software at http://www.census.gov/ftp/pub/ipc/www/imps.html).

CENVAR is a component of a statistical software system designed by the U.S. Bureau of the Census for processing, management, and analysis of complex survey data, the Integrated Microcomputer Processing System (IMPS). The program operates in a DOS environment but is screen oriented and

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

MEMBERSHIP APPLICATION FORM

I wish to enroll as a member of the International Association of Survey Statisticians.

Name : Mr./ Mrs./ Miss *First name :*

Date and place of birth:

Nationality :

Profession (if retired, please indicate) :

Title (professor, doctor) :

Business address :

Telephone and Extension :

Fax number :

Electronic mail :

Home Address :

Country :

Telephone N° :

I am a member of the ISI ☐

My mother tongue is (Please specify) :

My IASS correspondence should be sent to :

- *Business address* ☐
- *Home address* ☐

The official language of my IASS correspondence should be :

- *English* ☐
- *French* ☐

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

1. *Theory of surveys* ☐
2. *Collection of data* ☐
3. *Data processing* ☐
4. *Quality of data : errors of measurement, sampling errors, non–sampling errors* ☐
5. *Analysis of survey data* ☐
6. *Other (please notify)* ☐

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

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

I am aware that the membership dues have been fixed at 130 F. Francs (or the equivalent in other currencies) and that these dues have been reduced to 65 F. F. for nationals of developing countries.

Modes of payment

☐ *Cheque payable to theInternational Association of Survey Statisticians**

☐ *UNESCO coupons (for countries with nonconvertible currencies)*

☐ *Bank account of the I.A.S.S. 4586–32 with the Banque Nationale de Paris–Agence Didot, 67, boulevard Brune, 75014 Paris, France*

☐ *Postal account : CCP Paris n° 12 233 18 Z*

☐ *Other*

***Sent directly to : Secrétariat de l'AISE/IASS c/o Insee DR Aquitaine**
Att. Mme Claude OLIVIER, Bureau 136,
33, rue de Saget,
33076 Bordeaux Cedex,
France

Date: Signature:

IASS Short Course Registration Form

Check courses you will attend:

- ☐ A. Workshop on Survey Sampling in Developing Countries
- ☐ B. Variance Estimation in Complex Surveys
- ☐ C. Quality Management in Surveys (Jointly with International Association of Official Statisticians)

Check one:

- ☐ Developing Country
- ☐ Developed Country
- ☐ Student (If Student, please enclose evidence)

Check one:

☐ Payment enclosed, amount _____
(Make check payable to IASS Short Courses. You may use any convertible currency, US dollars preferred, using the exchange rate at the day of payment).

☐ Please send invoice for _____ (amount), for payment by 1 May 1997.

☐ Please register me provisionally, pending my search for financial support.

☐ The IASS is seeking financial support from several international agencies for participants from developing countries. If you are from a developing country and would like to be considered for such support please check here. (If checked, please enclose information about your current position and why you wish to attend the courses. Further information may be sought later.)

Accommodations

While attending the short course program, do you plan to stay in the hotel you will be in for the ISI session, or do you wish to reserve a separate local accommodation?

☐ ISI hotel ☐ Local hotel

If you plan to stay at your ISI hotel, you should indicate that on your ISI accommodation form. If you wish to reserve local accommodation, please complete the information below:

Hotel category: ☐ Single ☐ Double

Date of arrival: _____ Date of departure: _____

Number of persons (including you):_____

Other requirements:_____

_____

_____

Name:_____ Title:_____

Organization:_____

Mailing Address:_____

Fax number and E-mail address: _____

Please return registration forms to:

Professor Fritz Scheuren, IASS Scientific Secretary, Department of Statistics, The George Washington University, Washington, DC 20052 USA, Fax: 1-703-549-1119, E-mail: scheuren@aol.com

completely menu driven. Sample designs that can be accommodated include simple random sampling, stratified random element sampling, and multistage cluster samples with equal or unequal probabilities of selection. These sample designs are all addressed through the ultimate cluster sampling model. CENVAR uses Taylor series approximation for variance estimation, based on PC CARP software developed by Iowa State University (see below). Users must have a PC compatible microcomputer with 640K RAM and at least 10 Mb of hard disk space, DOS 3.2 or higher, and a printer capable of 132 characters per line. A math co-processor is highly recommended for Intel 286, 386, and 486SX processors. IMPS version 3.1 is easily downloaded from the WWW site, and installation is relatively straightforward. Only a subset of the IMPS needs to be installed to run CENVAR. Input data must be in ASCII format, and users must create an IMPS dictionary to read data. Blank fields are not acceptable and must be filled with zeroes before IMPS can read the data. Documentation is available in Word Perfect 5.1 format. CENVAR is not difficult to use, but users must be cautious about missing data; it appears to us in casual use that it will not accept missing values for any variables for a record. CENVAR produces sampling error estimates for means, proportions, and totals for the total sample as well as specified subclasses in a tabular layout. In addition to the standard error, 95% confidence interval limits, coefficients of variation, design effects, and unweighted sample sizes are given. We did experience some problems with the calculation of the design effects - users may want to double-check the figures.

CLUSTERS (World Fertility Survey; contact Vijay Verma, 105 Park Road, Teddington (Middlesex), TW11 OAW, United Kingdom; E-mail vjverma@essex.uk; no WWW access to software).

CLUSTERS is a stand alone program originally designed by the staff of the World Fertility Survey, and later updated and improved by Vijay Verma and Mick Price. It is now distributed for a nominal charge for handling by Vijay Verma. The program operates in a DOS environment and is keyword driven. Users must prepare in a file keyword commands in an ASCII format, and the location of that command file is passed to the program during execution. The principal sample design is a stratified multistage cluster sample, addressed through the ultimate cluster sampling model. CLUSTERS uses Taylor series approximation for variance estimation. Users must have a PC compatible microcomputer with 640K RAM and at least 2 Mb of hard disk space for the program, DOS 3.2 or higher, and a printer capable of 132 characters per line. A math co-processor is highly recommended for Intel 286, 386, and 486SX processors. The program comes with an extensive set of sample set ups and sample data sets for testing installation. CLUSTERS itself is easily installed, although some users may find execution a bit confusing until they learn how to specify the various required and optional input and output files. Input data must be in ASCII format, and users must create a dictionary in the command file to read data. Documentation is available in ASCII format. CLUSTERS produces sampling error estimates for means and proportions for the total sample as well as specified subclasses and subclass differences in a spreadsheet-like layout. In addition to the standard error, coefficients of variation, design effects, unweighted sample sizes, and intracluster homogeneity estimates are given.

Epi Info (U.S. Centers for Disease Control and Prevention; contact Andrew G. Dean, MD, Epidemiology Program Office, Mailstop C08, Centers for Disease Control and Prevention, Atlanta, GA 30333; E-mail AGD1@epo.em.cdc.gov or EpiInfo@cdc1.cdc.gov; WWW access to software http://www.cdc.gov/epo/epi/epi.html).

Epi Info is an epidemiological and statistical software system designed by the U.S. Centers for Disease Control and Prevention. It has features for processing, managing, and analyzing epidemiological data, including complex survey data (CSAMPLE component). The program operates in a DOS environment but is screen oriented. Documentation is available on-line in the program, and can be printed chapter-by-chapter. The basic sample design that can be accommodated is stratified multistage cluster sampling through the ultimate

cluster sampling model. Epi Info uses Taylor series approximation for variance estimation. Users must have a PC compatible microcomputer with 640K RAM and at least 10 Mb of hard disk space, DOS 3.2 or higher, and a printer capable of 132 characters per line. A math co-processor is highly recommended for Intel 286, 386, and 486SX processors. Epi Info is easily downloaded from the WWW site, but make sure you download a copy of the README.TXT file to get valuable installation information about installation on a PC or network. Installation is relatively straightforward, although it appears necessary for the user to create the needed subdirectories in advance. We loaded the entire Epi Info program even though we only needed a subset to estimate sampling errors. Input data can be from DBF, Lotus, or ASCII format. Users must create a questionnaire in Epi Info in the EPED component to read ASCII data; DBF and Lotus format entry are easier. The CSAMPLE component is entirely menu-driven. Epi Info produces sampling error estimates for means and proportions for the total sample as well as for subclasses specified in a two-way layout. The printed output includes only unweighted frequencies, weighted proportions or means, standard errors, 95% confidence interval limits, and design effects.

PC CARP (Iowa State University; contact Sandie Smith, Statistical Laboratory, 219 Snedecor Hall, Ames, IA 50011; E-mail sandie@iastate.edu; WWW access to software http://www.statlib.iastate.edu/survey/software/pccarp.html).

PC CARP is the PC version of the mainframe program SUPER CARP, Complex Analysis Regression Program, developed at Iowa State University. PC CARP can be used to estimate standard errors for means, proportions, quantiles, ratios, differences of ratios, and entries as well as test statistics for two way contingency tables. There are three companion programs which enlarge the range of analyses available: PC CARPL for logistic regression; POSTCARP for poststratified estimates of totals, ratios, and differences of ratios; and EV CARP for regression analysis, including under measurement error in the explanatory variables. The program operates in a DOS environment and is keyword oriented. The programs are designed to handle stratified multistage cluster samples with finite population corrections at up to two stages of selection. PC CARP uses Taylor series approximation for variance estimation. Users must have a PC compatible microcomputer with 640K RAM, and a math co-processor is highly recommended for Intel 286, 386, and 486SX processors. PC CARP and its companion programs may be purchased from the Statistical Laboratory at Iowa State University at $300. Input data can be in ASCII format, and users must create a dictionary to read data. Printed documentation is distributed with the software.

Stata (Stata Corporation; contact Stata Corporation, 702 University Drive East, College Station, TX 77840; E-mail stata@stata.com; WWW site http://www.stata.com).

Stata is a programmable statistical analysis software system which has recently begun to introduce commands which allow users to compute sampling error estimates for many statistics. The program is available for DOS and Windows environments and is keyword driven. Menus and help screens are available in the Windows version. We have only recently obtained preliminary documentation for the Stata commands for survey data analysis, and we are not sure what sample designs can be accommodated. Stata survey analysis commands use Taylor series approximation for variance estimation. Stata itself urges users to have a floating point processor on their PC, and Windows users are recommended to have 8 Mb of RAM and at least 4 Mb of hard disk space. The list price for Stata is $945 for commercial and $395 for academic users, and the survey analysis commands are included as part of the package. Since we have not yet purchased the software, we cannot comment on installation or use of the survey analysis commands, or input data format and documentation. The current survey analysis commands include svymean, svytotal, svyratio, and svyprop for means, totals, ratios, and proportions. The commands svyreg, svylogit, and svyprobt are available for the obvious regression, logistic regression, and probit analysis procedures. The commands svylc and svytest allow estimation of linear combinations

of parameters and hypothesis tests. The command svydes allows the user to describes the specific sample design and should be used prior to any of the above commands. There are plans to add survey analysis commands for estimating distribution functions and quantiles, contingency table analysis, missing data compensation, and other analyses.

SUDAAN (Research Triangle Institute; contact SUDAAN Product Coordinator, Statistical Software Center, Research Triangle Institute, 3040 Cornwallis Road, Research Triangle Park, NC 27709-2194; E-mail SUDAAN@rti.org; WWW site http://www.rti.org/patents/sudaan.html).

SUDAAN is a statistical software package for analysis of correlated data, including complex survey data. It provides facilities for estimation of a range of statistics and their associate sampling errors, including means, proportions, ratios, quantiles, cross-tabulations, odds ratios; linear, logistic, and proportional hazards regression models; and contingency table analysis. The program uses Taylor series approximations for variance estimation. It accommodates with and without replacement selection of first stage units, including components of variance, as well as simple random sampling and stratified element sampling designs. SUDAAN is available for PCs under MS DOS as well as for Windows, and prices vary depending on the type of firm, renewal status, and number of licenses. For example, the single license price for a new license PC version of SUDAAN 6.53 for commercial companies and government agencies is $995; the Windows version 7.0 is $1495. Current pricing lists are available directly from RTI. Although the most expensive of the software programs listed, users get a lot for their money. SUDAAN will read directly PC SAS data sets (but not the SAS for Windows data sets). The program is keyword driven, even in the Windows version, but the keyword syntax is very much like SAS. ASCII data sets can also be used in SUDAAN. Printed documentation is provided with each licensed copy of SUDAAN.

VPLX (U.S. Bureau of the Census; contact Robert E. Fay, U.S. Bureau of the Census, Room 3067, Bldg. 3, Washington, DC 20233-9001; E-mail rfay@census.gov; WWW access to software at http://www.census.gov/sdms/www/vwelcome.html).

VPLX is a stand alone program for variance estimation designed and used by the U.S. Bureau of the Census for complex survey data. The program operates in a DOS environment and is keyword driven. VPLX is primarily designed for stratified multistage cluster samples under the ultimate cluster sampling model. It uses repeated replication methods for variance estimation, including a random group, a jackknife, and a balanced repeated replication procedure. Users must have a PC compatible microcomputer with 8 Mb RAM and at least 3 Mb of hard disk space, and DOS 5.0 or higher. A math co-processor is highly recommended for Intel 286, 386, and 486SX processors. VPLX is easily downloaded from the WWW site, and installation is a matter of copying files to directories (there is no separate installation program). Input data must be in ASCII format, and users must create a dictionary to read data. Documentation is available from the Web site in Adobe format. An Adobe reader is also readily downloaded from the WWW. VPLX is more complicated to use than other keyword driven programs on this list, and requires, like CLUSTERS, program file be developed separately in an ASCII format for input to the program execution. It produces sampling error estimates for means, proportions, and totals for the total sample as well as specified subclasses.

WesVarPC (Westat, Incorporated; contact Westat, Inc., 1650 Research Blvd., Rockville, MD 20850-3129; E-mail WESVAR@westat.com; WWW access to software at http://www.westat.com/wesvarpc/index.html).

WesVarPC is a statistical software system designed by Westat, Inc. for analysis of complex survey data. The program operates in a Windows (3.1, 3.11, and 95) environment and completely menu driven. The primary sample design that can be accommodated is a stratified multistage cluster sample based on the ultimate cluster sampling model. WesVarPC uses repeated replication for variance estimation, including jackknife, balanced half sample, and

the Fay modification to the balanced half sample method. Users must have a PC compatible microcomputer with 4Mb RAM and at least 10.1 Mb of hard disk space, and Windows 3.1, 3.11, or 95. It is easily downloaded from the WWW site, and installation is probably the easiest among all the software listed using the installation software provided. Documentation is in an Adobe format, but instructions are provided on how to download the Adobe reader. The documentation is the easiest to read and best laid out among all the programs listed. Input data can be in ASCII format, or DBF, SPSS for Windows, SAS Transport, or PC SAS for DOS format. Input from any one of these formats is easy in the menu driven environment. WesVarPC requires that a new version of the data set be created in a special WesVarPC format. This requires the specification of replicates and, if poststratification is to be incorporated into the variance estimates, replicate weights. Users unfamiliar with these procedures may find the specifications slightly confusing. WesVarPC has facilities at present for contingency table analysis, regression, and logistic regression. There is an extensive menu driven system for creating new variables which extends the range of statistics that WesVarPC can be used for. Output is in a list format with one line per statistic in a simple format. The format is not suitable for publication, but it can be saved to a file for processing in a spreadsheet or other program.

Concluding Remarks

It is a bit difficult to summarize in briefly the features of each of these sampling error estimation programs. Some of these programs are have many components, and we may have not portrayed their features accurately. We certainly have not been able to give complete descriptions of the many features of each program. If there are errors, we apologize. They were inadvertent, and we would be happy to publish a revised description of any software description. Similarly, we are not sure we have been comprehensive in our catalogue. If there are other programs you are aware of that are available to survey analysis public that estimate sampling errors taking stratified multistage sampling into account, please let us know. Perhaps we can publish an updated list of such software in a future edition of the *Survey Statistician*.

Fortunately, our inaccuracies and omissions can, and probably will, be corrected in future issues of the *Survey Statistician*. We are in the process of inviting authors of the software to prepare descriptions of their software for the *Survey Statistician*. For example, in the next issue we hope to present an article by the authors of IMPS and its sampling error component CENVAR. Jim Lepkowski will serve as the editor for this feature in future issues.

There are other approaches to sampling error estimation possible besides these programs. For example, we have at the University of Michigan for some time used a primitive set of macros in SAS to estimate sampling errors for a wide variety of statistical analysis methods using repeated replication procedures. For extremely limited problems, some students have actually created spreadsheets in a few of the more popular spreadsheets to compute sampling error estimates from weighted cluster totals. Our goal here has been only to catalogue currently available software that any user could obtain access to.

Finally, there are bound to be new programs developed, and modifications and enhancements of the programs we have listed. For example, David Bellhouse at the University of Western Ontario produced the TREES sampling error estimation program several years ago. While he is not sure it will operate on current computing platforms, he is developing new software that uses computer algebra to derive formulae for variances of specified estimators. It may be useful to prepare a revised catalogue as undiscovered existing software, new software, or important modifications of existing software become available to us.

For further information, contact Jim Lepkowski at the Institute for Social Research, 426 Thompson Street, Ann Arbor, Michigan 48103, or by E-mail at jimlep@umich.edu.

References

Institute for Social Research (1984). *The OSIRIS Statistical Software System.* Ann Arbor, MI: Institute for Social Research.

Kalton, G. (1979). Ultimate cluster sampling. *Journal of the Royal Statistical Society, Series A* **142** (2) 210-222.

Kish, L., and Frankel, M. (1970). Balanced repeated replications for standard errors, *Journal of the American Statistical Association* **65**, 1071-1094.

Kish, L., and Frankel, M. (1974). Inference from complex samples (with discussion), *Journal of the Royal Statistical Society, Series B*, **36**, 1-37.

Lee, E.L., Forthofer, R.N., and Lorimor, R.J. (1989). *Analyzing Complex Survey Data*. Beverly Hills, CA: Sage Publications, Inc.

Lehtonen, R., and Pahkinen, E.J. (1995). *Practical Methods for Design and Analysis of Complex Surveys*. Chicester: John Wiley and Sons.

Skinner, C.J., Holt, D., and Smith, T.M.F. (1989). *Analysis of Complex Surveys*. Chicester: John Wiley and Sons.

.....

Question/Answer

The following two questions were raised by participants from Egypt and Turkey, respectively, during recent workshops on sampling. As always, we welcome alternative views on the issues raised, and also urge *The Survey Statistician* readers to raise questions of practical and common interest. Please send questions and observations on this column to: Professor Vijay Verma, University of Essex, Wivenhoe Park Colchester CO4 3SQ, United Kingdom, Fax: 44 1206 873151, E-mail vjverma@essex.ac.uk.

Q35.1. A national political opinion survey has been conducted based on a sample of blocks selected with PPS (probability proportional to populations size) at the first state. In order to survey a relatively rare population (unregistered female voters), we would like to expand the sample to cover substantially larger areas around each sample block. How can this be done?

As you subsequently clarified, the blocks used in the existing sample form subdivisions of polling districts, and the latter are considered to be suitable units for the expanded sample. In the existing sample each block (say a_{ij}) can be simply replaced by the polling district (say A_i) in which it lies. The resulting sample would be equivalent to the selection of whole districts at the first stage, with the probability of selection of a district equaling the sum of the selection probabilities of all the blocks contained within it: $P(A_i) = \Sigma P(a_{ij})$, summed over all blocks in the district in the sampling frame. If, occasionally, a district contains two blocks selected into the original sample, then it is considered as having been selected twice into the enlarged sample. For a district with more than two selections, a similar approach is used.

In the above example, the large units (polling district) constitute a set of fixed, nonoverlapping units in the population, and any originally selected unit (block) may be located anywhere within the boundaries of its district. While the above simple procedure suffices for an expansion of the sample in question, we may also note a more complicated situation in which the sample has to be expanded to cover larger areas, *but each demarcated so as to have an original area as its geographical center.* This has arisen, for instance in the Demographic and Health Surveys where the main survey of women in child-bearing ages is supplemented by a survey of family planning and health-care facilities covering extended areas around each of the original sample clusters. The supplementary survey has two objectives: (i) to survey facilities accessible to individual respondents, say located within 'x km' of each cluster in the original sample; and (ii) to obtain a national representative sample of different types of facilities themselves. For the first objective, the original individual respondents remain the units of analysis and no new sampling issues are involved: characteristics of the facilities associated with (accessible to) each original sample persons simply form additional variables in the analysis of the latter. However, the second objective, i.e. obtaining a representative sample of *facilities* by demarcating larger areas (A_i) around each original cluster (a_i) as the center, requires special treatment. Clearly, the A_i's now constitute a set of overlapping units in the population. The issue is to identify appropriate (valid and practical) rules, for the demarcation of the larger areas such that the

resulting A_i's constitute a proper sample of areas with known probabilities of selection.

Space does not permit a detailed discussion of this problem in the Q/A Column this time. Nevertheless, it will be useful to indicate one possible solution, which is on the following lines. I can deal with it in more detail in the next issue, or submit a separate technical note on the issue to *The Survey Statistician*. In the meantime, *The Survey Statistician* readers are invited to propose their own solutions.

A solution. If the larger areas around the original sample clusters are demarcated such that (i) each constitutes a circle of fixed size with the original cluster as the center (such as some specified distance around it); and (ii) the sum of probabilities of the clusters contained with each circle is also constant, then we obtain a valid sample of the areas defined by the circles so demarcated, each selected with probability.

If the original clusters have been selected with probabilities proportional to population size (as is often the case), the above two conditions can be satisfied simultaneously only if the population density is uniform. However, in practice this is often not too restrictive a requirement because the two conditions need to be satisfied only 'locally' i.e. among the A_i's which overlap. It is usually possible, for instance, to divide the country into strata each reasonably uniform in population density and then to demarcate circles of some appropriately fixed size around the original sample clusters within each stratum. The size can vary from one stratum to another (such as inverse proportion to its average population density). Also, acceptable compromises can usually be found in practical application of the procedure sketched above.

Q35.2 In a six-monthly survey of establishments and earnings, the sample of establishments has been selected by dividing a list frame into a large number (nearly 1,000) strata by establishment size, industry (2 digit TSTC) and location (region). Despite the substantial overall sample size, sample sizes per stratum are very small - around 6 units per stratum on the average. Within each stratum, establishments were selected at a uniform rate, and we estimate the stratum total by inflating the sampling aggregate by the factor (N/n), where N is the number of units in the frame and n the number successfully enumerated. Can we improve on this estimation procedure? We have no external source of information for this purpose. Even, worse, we are unable to update the frame in-between full censuses several years apart. Also, it is not possible in practice to follow-up establishments which move.

This is a description of a problematic situation which I find remarkably common, especially in developing countries - though I hope that you are not being unduly pessimist, i.e., in your description and about the possibilities for improvement.

A number of issues relating to the problem were discussed earlier in this Column. Please refer to Q/A 33.1 of December 1995. A few additional points can perhaps be made usefully.

In a sense, your estimation procedures 'compensates' for nonresponse since in the inflation factor, (N/n), the denominator n is the number of actually enumerated rather than the originally selected units. However, on examination of the information subsequently provided one finds that, despite the large number of strata constructed, categories by establishment size (as measured by the number of employees at the time the frame was constructed) are generally quite broad. (Most of the detail in the stratification comes from the fine control by industry, which is quite common in such surveys because of the particular reporting requirements). Normally, the various quantities to be estimated from the survey (such as output or current employment) tend to be highly correlated with the original size, and therefore a ratio-expansion of the form (E/e), in place of the currently used factor (N/n), is likely to be appreciably more precise. Here E refers to the total measure of size (employment) of establishments in the frame for a particular stratum, and e the same measure summed for the establishments successfully enumerated in the survey.

In Q/A 33.1, it was suggested that in the absence of any possibility to include new establishments in the frame, the problem of the frame incompleteness may be reduced by permitting *substitution for closures* found in the sample. This is based on the

expectation that these closures compensate for at least a part of the newly created units entering the population which are not reflected in the frame. In so far as movements of the units cannot be followed-up and are not reflected in the frame for subsequent survey rounds, they may be treated in the same way as closures.

Given the very detailed stratification involved in the survey, it is possible that in certain instances individual strata are found to be too small to contain enough units to provide the required substitutes. In such cases, some merging (collapsing) of the original strata may be necessary for the purpose.

.....

Country Reports

AUSTRALIA (from Susan Linacre)

The Australian Bureau of Statistics recently ran a **Survey of Aspects of Literacy,** based on a methodology developed by the United States' Educational Testing Service and Statistics Canada, and adapted for use in several other countries as part of an International Study of Literacy coordinated by Statistics Canada. The study has involved similar surveys in Belgium, Canada, Germany, Ireland, the Netherlands, New Zealand, Poland, Sweden, Switzerland, the United Kingdom and the United States.

The survey provides a direct, objective assessment of functional literacy skills using examples of approximately 100 commonplace tasks of varying difficulty drawn from a range of areas. To minimize provider load while covering a variety of tasks, a balanced incomplete block design was used to allocate sub samples of the tasks to each respondent. Skill levels on a scale of 1-5 will be derived from the raw scores on the tasks.

A large set of background information is also collected for each respondent, including: socio-economic characteristics; educational attainment; labor force experience and status; parents' educational attainment and occupation; respondent's perception of their English literacy skills; respondent's literacy practices at work and in everyday life and barriers to English literacy training. First results from the Australian survey are scheduled for release in December 1996. (Contact Pam Lee, Fax 61 6 251 5486; E-mail pam.lee@abs.telememo.au.)

The **Survey of Employment and Unemployment Patterns** is a new longitudinal survey being conducted by the ABS. It aims to provide information on the dynamics of the labor market, and to assist in the assessment of the impact of labor market assistance initiatives in alleviating the extent of 'joblessness' in the Australian economy.

An initial panel of respondents, established in the first half of 1995, comprises three sub-panels designed to facilitate policy relevant analysis. Respondents will be interviewed again in 1996 and 1997 using CAI. The survey will yield an extensive set of data centered around episodal information relating to respondents' labor market activities and their participation in Labor Market Programs. With the respondent's consent, the direct data collected will be supplemented with information from the administrative systems of the Department of Employment, Education, Training and Youth Affairs and the Department of Social Security.

Initial results from the survey, covering the period September 1994 to September 1995 will become available in the second half of 1996. (Contact Mel Butler, Fax 61 6 251 6571; E-mail mel.butler@abs.telememo.au).

The second in the series of 5 yearly **National Health Surveys** was conducted by the ABS in 1995. These surveys provide national benchmarks on a range of health and related topics, and enable trends in key aspects of health to be monitored over time. The 1995 survey, differed from previous surveys in that, while a common core set of data items was collected for all respondents, a split sample was used for some items to reduce respondent load. As well the sample was supplemented by additional indigenous sample, and a review of the quality of data collected for the indigenous population is currently being undertaken. (Contact: Mike Langan, Tel.: 61 6 252 6403, Fax: 61 6 253 1673).

The **National Nutrition Survey** (NNS) is a joint project between the Department of Health and Family Services and the ABS. It was linked with the National Health Survey (NHS) with approximately thirty per cent of persons who had completed the NHS being randomly selected to take part in the

Nutrition survey. The data will enable health planners to study issues such as the link between diet and health; risks of nutritional deficiencies; variations in eating patterns in different parts of the country; and changes in eating patterns as people grow older.

Respondents provided very detailed data on all food and drinks consumed during the previous 24 hours; had their blood pressure, height, weight, waist and hip measurements taken; answered a range of food related questions; and returned a mail back questionnaire on their average frequency of food intake over the previous 12 months. (Contact: Marelle Rawson, ABS, Tel.: 61 6 252 7995, Fax: 61 6 253 1673).

The **Children's Immunization and Health Screening Survey** was conducted in April 1995 as part of the ABS monthly population survey program. The survey collected information on immunization coverage of children (relative to the National Health and Medical Research Council's recommended schedule) and on levels of sight and hearing tests and dental visits among children. Results of the survey are available. (Contact: Brian Richings, ABS, Tel.: 61 6 252 5786, Fax: 61 6 253 1673).

CANADA (from Gordon Brackstone)

The monthly **Canadian Labour Force Survey** (LFS) is in the final phase of a major redesign which has included changes in four major areas. In the first phase, computer assisted interviewing was tested as a replacement for traditional paper and pencil interviewing. The computer application, containing an unaltered version of the paper questionnaire, was introduced at the end of 1993. In the next phase a new sample design was introduced between October 1994 and March 1995. The new design is simpler, more flexible and more efficient than the previous one. To meet a budgetary reduction, the LFS sample was reduced by 11 percent in July 1995. The sample reduction was done in such a way that the quality of provincial and national estimates remains comparable to the quality prior to the introduction of the new sample design. In the third and final phase of the redesign, a new questionnaire and new computer systems are being introduced simultaneously. The old questionnaire is essentially the same as the one introduced in the 1970s. The new questionnaire incorporates improvements to the wording and sequencing of questions to better deal with the modern labor force. It also includes new content, including questions on earnings and union membership. Some of the computer systems used by the LFS are out of date and difficult to maintain. The new systems are based on state-of-the-art Unix-based software and technology. Both the new systems and the new questionnaire, which have been tested extensively, are being introduced in the last quarter of 1996.

More detailed information on the new questionnaire is available in English and French on the World Wide Web at www.statcan.ca/Documents/English/Subjects/Labour/Survey/labour.htm. For additional information, contact Jack Gambino at (613) 951-0334, Fax: (613) 951-3100 or E-mail: gambino@statcan.ca.

The **Workplace and Employee Survey** is a new Statistics Canada undertaking designed to provide an integrated view of the activities of employers and employees. There is a need to link workplace practices and employee outcomes and characteristics. This is important as there have been fundamental changes in the labor market, and in workplaces (establishments) during the eighties and nineties. These changes are not well understood because of the lack of appropriate data. Some of these changes are affecting workers in the labor market: examples include stagnation of average wages, rising inequality of earnings, apparent slowdowns in full-time job creation and the development of core and temporary work forces. There is also a perception of fundamental changes within the workplace: these include implementation of information technologies, innovation and firm performance, increasing competition and the search for more "flexible" organizations. Existing data sources originating with household surveys and establishment data sources address these issues independently.

A pilot test was conducted in 1995-1996 to assess the feasibility of conducting this survey on a production basis. The Workplace and Employee Survey pilot was designed as an employer component, based on a two-stage sample, and a third stage sample drawn for the employee component. The pilot survey data are being analyzed, and a yearly longitudinal production survey is scheduled to start in 1997-1998. For

more information contact: Alain Théberge at (613) 951-1466 or E-mail: THEBERG@STATCAN.CA.

CHINA (from Langhui Huang)

Along with the establishment of the social market-oriented economy in China, the State Statistical Bureau (SSB) has formulated a statistical reform strategy that includes periodic censuses as a basis, and regular sample surveys as a principal method of collection, supplemented by the use of estimating methods. Much progress has been made in 1996.

The Third National Industry Census (TNIC) is in progress, with the base year of 1995. The scope of TNIC includes all industrial enterprises and affiliated industrial units engaged in mining, manufacturing and production, and the distribution of electricity, gas and water. Large and medium sized enterprises receive a long questionnaire, while small enterprises receive a short questionnaire. Village and private enterprises with annual income less than one million Yuan (RMB) provide data only on some basic indicators.

Another census being conducted this year is the First National Statistical Unit Census (FNSUC). It covers corporate enterprises and establishment units using SNA definitions.

After a year's experiment, the nationwide monthly sample survey of wholesale, retail and catering enterprises has been officially brought into the National Statistical System. This represents a breakthrough point in the transition from the traditional reporting system to a survey system.

Since September 1996, the monthly sample survey of industrial enterprises has been carried out on an experimental basis.

For more information, please contact: Mr. Langhui Huang, Director, Enterprise Survey Organization (ESO), State Statistical Bureau (SSB), Beijing 100826, P.R. China. Tel.: 86-10-63401425, Fax: 86-10-68576354.

ITALY (from Claudio Quintano)

The Italian National Institute of Statistics (ISTAT) is undertaking some innovative activities in the area of enterprise registers.

In 1996 the Statistical Archives of Active Enterprises (ASIA) and of local units will be completed, and produced in a prototype form on a provincial basis. ASIA is based on past censuses updated through linkage with microdata on active enterprises contained in large administrative databases. The innovative aspects of the ASIA Project include the conceptual, organizational and informatic architecture of the system as well as purely methodological aspects.

Within the enterprise statistical system, the new archive will produce:

1. integrated records of administrative data referring to the same unit and derived from various administrative sources that will be returned to the furnishing bodies;

2. territorial statistical data updated yearly;

3. lists of individual enterprises updated during the year for each territory.

The last two "products" will permit more efficient design of surveys, as well as greater statistical utilization of administrative data on enterprises. The statistical units being observed through ASIA will be georeferenced via the geocoding system already used by ISTAT.

By the fall of 1996, new surveys will be introduced on some service sectors (informatics, quality certification, publicity, marketing, communication) that will enable a highly articulated information base on third party services in Italy.

Statistics associated with consumer prices, employment and wages in large enterprises, and sales indexes in wholesale trade will be referenced to 1995. In these last two cases, as well as producing the indexes, innovations will be introduced in the information content of the questionnaires in response to economic analysis needs and the statistical structure of the surveys. With regard to consumer prices, a system geared to price indexes is being implemented to take account of both the different means of distribution (traditional, modern, *etc.*) and the different patterns of consumption, which are also linked to the family's life cycle. Another linked information product consists of the new system of indexes on Foreign Trade, base year 1993, which will enable the

monthly production of data on values, prices and quantities and commercial flows, linked by sector and geographic area, as well as sub-indexes of particular relevance to economic analysis (*e.g.*, indexes of the cost of imported raw materials).

Experimental activities on using administrative data will be concluded in 1996 (particularly those relevant to employment, hours worked and compensation). In particular this entails integrating monthly data regarding large enterprises which have been gathered directly by ISTAT with the results of surveys of small and medium sized firms obtained from the National Social Security Administration.

Under methodological aspects, experiments have been conducted which are resulting in innovative methods being already incorporated into the production processes of some structural surveys. These include techniques for missing responses that are relevant to enterprises data, the treatment of social transformation, and links between the micro and macro levels in the economic accounts of enterprises.

For further information, please contact Roberto Monducci, DCII/A, ISTAT, Via Cesare Balbo 16 - 00100 Rome (Italy), tel. 39-0-6-46732455.

JAPAN (from Chikio Hayashi)

The Japan Broadcasting Corporation (NHK) introduced a new pre-code method for its 1995 **Time Use Survey.** The previous after-code method was replaced for the following reasons: after-coding required an enormous amount of time and labor given that NHK's time use surveys cover more than 10,000 people; NHK's programming and program production divisions strongly requested a shortening of the time taken to process results and produce reports; and, due to social changes in Japan, it had become increasingly difficult to secure temporary workers as coders.

Based on the observations of a pretest, we conducted in 1995 a main survey using the pre-code method and also a supplementary time-series survey using the after-code method.

From the results of these surveys, we can point to the following three characteristics of time use in Japan for 1995:

1. With the spread of the five-day work week, lifestyle on Saturdays has come to resemble that on holidays.

2. With the widespread adoption of the five-day work week and the partial introduction of the five-day school week over the past five years, a growing number of Japanese are now at home on Saturdays. As a result, activities on Saturdays are becoming like those on holidays. To be specific, more time is spent on TV viewing, and time for household chores has increased. The time for getting up tends to be later and sleep time is longer.

3. TV-viewing time showed a major increase. The biggest change has been in the length of TV-viewing time. Regardless of the day of the week, age group, or time slot, the number of TV-viewers has been on the rise, with TV-viewing time reaching its highest point ever in 1995. The background factors of this change include the decline in time for paid work, school work, and household chores, mainly on Saturdays; rising evaluation of TV-viewing; and a series of major incidents and accidents that attracted large TV audiences.

4. The gender gap in time use continues. The length of time that Japanese women spend on household chores keeps declining as a result of the tendency towards nuclear families, and the widespread use of electric household appliances and processed foods that help save time. On the other hand, the time men spend on household chores has gradually been on the rise, reflecting men's changing views about household participation. This has somewhat narrowed the gender gap regarding household chores, compared with previous years, but there is still a considerable gap between men and women. Furthermore, men tend to sleep longer than women. In short, gender characteristics regarding Japanese time use have not changed fundamentally.

The survey outline is as follows:

	Main Survey	Supplementary Survey
Survey period	Oct. 12-25, 1995	Oct. 14-15, 19-22, 24-25, 1995
Sample size	25,200 aged 10 & over	7,200 aged 10 & over
Survey method	Pre-coded questionnaire	After-coded questionnaire
Responses	19,181 (76.1%)	5,318 (73.9%)

For more information, please contact Tetsuo Makita, Japan Broadcasting Corporation, Public Opinion Research Institute, 2-1-1 Atago, Minato-ku, Tokyo 105, Japan. Tel.: 81-3-5400-6861, Fax: 81-3-3438-4375.

PERU (from Leonor Laguna)

In 1995, the Instituto Nacional de Estadistica e Informatica (INEI) started a National Household Survey carried out quarterly through a master sample. The sample size is about 20,000 housing units at national level. The samples are independent in each quarter and in each department (Peru is politically divided into 24 departments). The results will be presented by department, rural and urban areas, department capitals and geographical regions (costa, sierra, selva).

The sampling design is self-weighting, multistage (4 stages in the urban areas and 3 in the rural areas).

The subjects of study were organized and carried out through the quarters in the following way:

95-1 Employment and Income, Health and Education

95-2 Employment and Income, Health and Transport and Communications

95-3 Employment and Income and Evaluation of the scope of social programs undertaken by the Government

95-4 Employment and Income, Measurement of Level of Life: Health, Social Programs and Education (specially scholar absenteeism and desertion).

In 1996, the following innovations were made:

96-1 For Employment and Income a panel sample with partial rotation was implemented

96-2 One fourth of the panel was changed

96-3 A special survey on levels of employment was undertaken in agreement with the Labor Ministry.

In the second week of September representatives of the organization for the Improvement of Quality of Life Surveys in Latin America and the Caribbean (MECOVI) and INEI had meetings to agree on an action plan in Peru for the period of 1997-2000.

PHILIPPINES (from Gervacio G. Selda, Jr.)

Release of the 1995 Population Census Results. The mid-decade or 1995 count of the Philippine population was officially released last August 14, 1996 through a presidential proclamation. Based on the results of the count, the total population of the Philippines as of September 1, 1995 was 68,614,162. The census also revealed that: (a) the country's population grew at a rate of 2.32 percent annually in the first half of the nineties as against 2.35 percent annually from 1980 to 1990; (b) average size of Filipino household is 5.07 persons as compared to 5.31 persons in 1990; and (c) the Philippine population is expected to double in less than 30 years based on 2.32 percent annual growth rate. The conduct of the population census in the Philippine is a major activity of the National Statistics Office. For more information, contact Elizabeth Go, National Statistical Office, Manila, Philippines, E-mail: ego@census.gov.ph or Fax: (632) 716-0404.

The Philippine government recently adopted a system of **designated statistics** to facilitate the generation of the most critical and essential statistics required by administrators, planners and policy makers in the government and private sectors. In adopting the system, twenty-eight (28) censuses and surveys, seventeen (17) administrative data systems, four (4) derived data systems and six (6) statistical indicators were designated. The list will be updated from time to time as the need arises. In general, the system of designation entails the following: (a) regular updating of the design for generating the designated statistics; (b) setting aside government funds to finance the budgetary requirements of the designated activities; and (c) dissemination or release of data from the various designated activities and systems in accordance with the time frame. The National Statistical Coordination Board was tasked to handle the maintenance and updating of the system. For more information, contact Francisco K. Mallion, National Statistical Coordination Board, Makati, Philippines, E-mail: nscbasg@mozcom.com or Fax: (632) 890-9408.

The **Survey of Children 5-17 Years Old** was undertaken in 1995 by the National Statistics Office in close collaboration with the Bureau of Labor Employment Statistics of the Department of Labor and Employment as a component of ILO's International Programme for the Elimination of Child Labor.

The study was a rider to the July 1995 Labor Force Survey (LFS). As such, it adopted the LFS stratified two-stage sampling design. The urban and rural areas of each province were the principal domains of the survey. Barangays were considered as the primary sampling units (PSUs) and households as the secondary sampling units (SSUs). Barangays in each domain were arranged by population size in descending order and grouped into strata of approximately equal sizes. Household as the secondary sampling units were selected systematically with a random start. The sample included 2,100 barangays and approximately 25,500 households nationwide. All children aged 5 to 17 years old who were found to have worked at any time during the past twelve months at the time of the survey were interviewed.

Among the major findings of the study were: (a) 3.67 million children or 3 in every 20 children have worked in the past; (b) males comprised two-thirds of the working children; (c) two-thirds of working children come from rural areas; (d) the majority or 56.0% of the working children are in farming; (e) 2.8% of the working children did not complete any grade level at all; (f) of those who entered school, 17.7% were elementary graduates, 7.1% graduated high school and 1.5% reached college; (g) about 60.0% of the working children are exposed to a hazardous environment; (h) about 0.57 million working children expressed the need for improved working conditions; and (i) one out of 100 working children suffers from work-related injuries/illness. For more information, contact Elizabeth Go, National Statistics Office, Manila, Philippines, E-mail: nsohsd@i-manila.com.ph or Fax: (632) 716-0404.

The study to **redefine urban and rural areas** in the Philippines using statistical methods was completed already. The research was a joint undertaking of the National Statistics Office and the Statistical Research and Training Center with the UNFPA providing the research fund. In this study, multivariate analysis was applied to 1990 Census of Population and Housing data to determine the urban criteria. Cluster and exploratory data analysis were applied to 22 variables or major census information items. The aim of the undertaking was to come up with an urban redefinition which is simple and operational and which can be applied using regular census results without the need for more data gathering. Based on the proposed definition, urban population in the Philippines was estimated at 40.2 percent in 1990. For more information, contact Teresita E. Valerio, Statistical Research and Training Center, Quezon City, Philippines, Telefax: (632) 921-7485.

REPUBLIC OF KOREA (from Seung-Kon Lim)

The 15th Korean **Population and Housing Census** was conducted in November 1995 using 126,000 enumerators. The Census questionnaires were processed using Optical Mark Reader sheets for quick data processing as in the previous census in 1990.

The preliminary report was released in March 1996. The report, which was based on the summary sheets prepared by the census enumerators,

contains the number of persons, households and housing units by administrative districts.

In addition, the Advance Report, based on a 2% sample of households across the whole country, was released in July 1996 to meet urgent demands for estimates of the basic characteristics of population, households and housing units for the whole country and by province.

The major features of the results are as follows:

- The results show that 44,551,000 persons were counted as of November 1995;
- Among the population aged 15 years and older, 61.2% persons were married, 30.2% were unmarried, 7.5% were widowed, 1.1% were divorced;
- The total number of households counted was 12,961,000, an increase of 14.1% compared to the 1990 census;
- The average household size was 3.3 persons, compared to 3.7 persons in 1990;
- Households composed of two generations accounted for 62.8% of all households; and
- The total number of housing units counted was 9,216,000, an increase of 28.7% compared to the previous census.

The final report of the census will be released in December 1996.

For further information, contact Mr. Jang-Soo Yang, Director of the Population Statistics Division, National Statistical Office, no. 647-15, Yoksam-dong, Kangnam-gu, Seoul, Korea. Tel.: 82-2-222-1861, Fax 82-2-538-6348.

.....

SRMSNET

The Survey Research Methods Section (SRMS) of the American Statistical Association now has a listserv called SRMSNET. The main idea of a listserv is that subscribers post their message and it is then electronically mailed to all the subscribers. The mail will appear as any other mail message you now receive.

The SRMS has decided that, for the time being at least, it would allow anyone to subscribe, whether or not they are Section members. In this way, we expect that a broad community of survey statisticians will participate in the discussions. It is also a very timely way to get out interesting news, such as conferences and job openings.

To subscribe, simply send a message to listserv@umdd.umd.edu and, in the body of the message, type:

subscribe SRMSNET John Doe
(substitute your own name for John Doe).

Within 5-10 minutes you should receive confirmation of your subscription and instructions on how to post or get information.

And, what do you get for your efforts? In addition to a preview of the SRMS Newsletter, announcements about upcoming conferences and seminars, and job postings, some new software was offered free of charge as a means of obtaining feedback on it during the final production stages. Subscribers have also inquired about where to locate information on a particular topic and sought clarification on the technical meaning of statistical terms. A question about good books on pps sampling led to some lively exchanges and another inquiry about an inexpensive alternative software for data analysis generated much discussion.

David Binder, 1996 Chair
Survey Research Methods Section
American Statistical Association

David A. Binder, Director
Business Survey Methods Division
11-A R.H. Coats Building
Statistics Canada
Ottawa, Ontario, CANADA K1A 0T6
binddav@statcan.ca
TEL: (613) 951-0980 (Office)

.....

IASS Program for the 1999 Session, Finland: What Topics Would You Like?
Susan Linacre

The program committee for the 1999 Session in Helsinki has been convened, and is seeking input from Association members on topics for the IASS share of the Meeting. There is expected to be scope for around 10-12 topics with a couple of reserves, given options for joint topics with other sections. Suggestions are welcome on any aspects of survey statistics, with preference generally given to topics of wide and practical interest. The topics could be either theoretical or applied, with an emphasis on new developments encouraged. Topics addressing problems faced by developing countries, as well as those of developed countries, are encouraged.

Suggestions may be either for a general area of interest for a session as a whole, or a more specific topic suitable for an individual paper. Possible names (and contact information) for organizers, speakers and discussants are also encouraged, as well as any indications of areas of new work in the topic field or new applications related to the topic.

While you are strongly encouraged to put forward new topic suggestions, some suggestions have already been put forward and, you are also invited to provide any comments on these topics. In particular there may be topics here for which you would like to express support. The list of suggestions to date is given below.

In providing suggested topics for the Session or comments on the topics below, please ensure you provide clear contact information (address, telephone or fax number, E-mail) for further correspondence. Suggestions and comments can be made to any member of the committee. The committee members are:

Susan Linacre, Australian Bureau of Statistics
Fax: 61 6 253 1093,
E-mail: sisd.exec@abs.telememo.au

Anders Christianson, Statistics Sweden
Fax: 46 8 783 4599

Sid David, Asian Development Bank, Manila
Fax: 63 2 636 2444
E-mail: ipdavid@mail.asiandevbank.org

Seppo Laaksonen, Statistics Finland
Fax: 358 0 1734 2474
E-mail: Seppo.Laaksonen@stat.fi

David Morganstein, WESTAT Inc., USA
Fax: 301 294 2034
E-mail: morgand1@westat.com

Danny Pfeffermann, The Hebrew University of Jerusalem, Israel
Fax: 97 2232 2545
E-mail: msdanny@olive.mscc.huji.ac.il

Fritz Scheuren, George Washington University, USA
Fax: 703 549 1119
E-mail: Scheuren@aol.com

Vijay K Verma, ESRC Research Centre, University of Essex, UK
Fax: 44 1206 873151
E-mail: vjverma@essex.ac.uk.

Topics Already Suggested

At this early stage, some of the topics suggested are very broad, while others are quite specific. In addition, there is some overlap. Comment on any aspects of these topics is sought. The topics are:

1. Understanding how clients actually use survey data, and the implications for sample design, data analysis, and data dissemination.

2. Presentation of information on the quality of statistics to users, and how to make it more useful.

3. Practical issues in data validation in surveys, including post enumeration surveys.

4. Evaluation of "short-cut" procedures *e.g.*, limited coverage, non probability sampling methods, proxy interviews.

5. Measuring the effects on time series of the introduction of new methodologies.

6. Issues of common interest to business and household based survey statisticians.

7. Coordination of samples and managing respondent load.

8. Timely reflection of changes in the small business population in business surveys.

9. Opportunities and pitfalls in using EDI for data collection.

10. CAI some years on: What have we learned of the opportunities and pitfalls?

11. Lowering the cost of data capture through innovative technologies, data reuse and data mining.

12. Statistics with non precise data.

13. The use of graphics in complex surveys.

14. Procedures, problems and practices in weighting sample data.

15. Sampling over time: cumulation rotation, rolling samples, longitudinal panel surveys.

16. Small domain estimates: an overview and recent developments.

17. Seasonal adjustment versus trend estimation - what should survey designers focus on.

18. Maximum likelihood estimation using complex survey data.

19. Use of multilevel models in survey estimation and analysis.

20. Surveying the informal sector.

21. Planning for the year 2000 Censuses of Population and Housing, and of Agriculture in developing countries, and the use of census information in survey designs for the following decade.

22. Use of Internet to link developing countries and international agencies, and issues in handling and sharing large databases in the Internet.

23. Strengthening of statistical systems in transition countries with an emphasis on Asian transition countries, and some comparisons of the issues with those of transition countries in Eastern Europe.

.....

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

The Survey Statistician

No. 35
December 1996

J. Michael Brick ..... Editor

Section Editors
Gordon Brackstone .....Country Reports
Vijay Verma ..... Question/Answer
James Lepkowski .....Software Review
Ann Marie Vespa-Leyder.....Circulation
Angelia Murphy.....Production

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

Secrétariat de l'AISE/IASS
c/o INSEE DR Aquitaine
Att. Mme Claude Olivier, Bureau 136
33, rue de Saget,
33076 Bordeaux Cedex, France

Comments on the contents or suggestions for articles in *The Survey Statistician* should be sent via E-mail to BrickM1@Westat.com or mailed to:

J. Michael Brick
Westat, Inc.
1650 Research Blvd., Room 490
Rockville, MD 20850 USA

.....

Short Course Announcement

The International Association of Survey Statisticians (IASS) is sponsoring a program of short courses to take place before the 51st Session of the International Statistical Institute. These courses will be led by international experts in their fields and will offer practitioners, researchers and students in

statistics and survey methodology, the opportunity to improve their skills and to discuss issues of common interest. Two of the courses will be aimed especially at statisticians working with surveys in developing countries. The course on Quality Management should be of interest to participants from other countries. All courses will be presented in English and participants should have the ability to work in this language.

The courses will be held in Istanbul, at or near the location where the main ISI Sessions will take place. As participants, you may choose either to extend your stay in the accommodations you arranged for the ISI Session, or you may stay in a local accommodation, arranged specifically for these short courses, by completing the accommodation section of the registration form.

Short Courses

A. Workshop on Survey Sampling in Developing Countries, 15-16 August, 1997. Presented by:

Graham Kalton, Westat Inc., Rockville, Maryland, USA

Colm O'Muircheartaigh, London School of Economics, UK

The emphasis of the workshop will be on practical aspects of survey sampling in developing countries. Examples from demographic and household surveys and from agricultural surveys will be drawn upon. The participants should have experience in the planning, implementation, evaluation or analysis of surveys. The workshop will be of most benefit to those who are involved in survey practice in developing countries.

Topics will include:

1. Frame construction;
2. Sampling methods, including stratification, multi-stage sampling and area sampling;
3. Maintaining control of sampling processes; and
4. Sampling weights and an introduction to variance estimation.

B. Variance Estimation in Complex Surveys, 17-18 August 1997. Presented by:

Wayne Fuller, Iowa State University, Ames, Iowa, USA

Kirk Wolter, National Opinion Research Center, Chicago, USA

F. Jay Breidt, Iowa State University, Ames, Iowa, USA

The purpose of this course is to provide training in variance estimation in complex surveys for survey statisticians, especially for those from developing countries.

The course will cover methods of estimating variances for statistics such as means, proportions, ratios, regression coefficients and statistics arising in the analysis of two-way contingency tables. The use of the program PC CARP for computing variances of statistics from complex sample designs on a personal computer will be demonstrated, and instruction will be given in the use of the program. About one half of the course will be devoted to methodological considerations and one half to instruction in the use of PC CARP.

Each registered participant will receive the program, PC CARP, the PC CARP manual and a complimentary copy of the course text:

K. M. Wolter (1985) *Introduction to Variance Estimation.* Springer-Verlag, New York, USA.

C. Quality Management in Surveys (Jointly with International Association of Official Statisticians), 18 August 1997. Presented by:

David Marker, Westat, Inc., Rockville, Maryland, USA

David Morgenstein, Westat, Inc., Rockville, Maryland, USA

Statistical agencies have always been concerned with product quality and the reduction of variation. This seminar presents a systematic approach to

process quality improvement throughout an agency, viewing the agency as a system of interlocking processes that produce a final product-- a census or a survey; data registries or publications.

Modern quality improvement contrasts with traditional methods that focus almost entirely on measuring the quality of the survey product with little recognition that better control of the processes improve the result.

In this seminar, the presenters will describe how statistical agencies can systematically identify and better control processes that produce statistical products and services. Examples will be presented from statistical offices in Sweden, Finland, and the United States. Participants should include top managers and others concerned with improving the quality of their statistical agency.

Time Schedule

Friday, 15 August 1997 (full day)	A		
Saturday, 16 August 1997 (full day)	A		
Sunday, 17 August 1997 (full day)		B	
Monday, 18 August 1997 (a.m. only)		B	C

Participants may attend course A followed by B, or course A followed by C, or else any single course alone.

Registration

Registration fees are as follows:

	Developing countries	Developed countries	Students
One course	$300	$400	$150
Two courses	$400	$500	$250

All prices are quoted in US dollars.

Applicants should return the completed registration form by 1 May 1997, in order that the organizers can assess the demand for courses. If there is insufficient demand, a course may be canceled. The registration form is enclosed with this issue.

.....

In Other Journals

Journal of Official Statistics

An International Review Published by Statistics Sweden

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

Contents Volume 12, Number 1, 1996

Contents Volume 12, Number 2, 1996

All inquires about submissions and subscriptions should be directed to the Chief Editor:
Lars Lyberg, R&D Department, Statistics Sweden, S - 115 81 Stockholm, Sweden.