International Statistical Institute

The survey statistician
nr 23

International Association
of Survey Statisticians

June 1990

C O N T E N T S

* * * *

I. NEWS FROM THE ASSOCIATION/GENERAL INFORMATION

1.1 Dues

The dues amount to F.F. 130 (F.F. 65 for members from developing countries). Any convertible currency is accepted, at the rate of the day of payment.

1.2 About the Survey Statistician

The IASS Council is initiating the process of reviewing the presentation and contents of the "Survey Statistician". We are anxious to hear from members about their views.

We would welcome comments on the "Survey Statistician" as it now exists with particular reference to how it serves members' needs. Are there sections which are of no interest, or others which should be retained ? Are there items of information or services which the "Survey Statistician" might fulfill but which are not covered at present ? If you have found items in recent editions of particular value or of no value at all then please let us know. Both positive and negative comments are equally valuable.

Please send all comments to :

Professor D. Holt
Department of Social Statistics
Southampton University
Southampton SO9 5NH
England.

II. CAIRO SESSION - 1991

List of the organizers of the IASS invited-paper sessions

Session 15. Survey Sampling in Marketing

Mr. William J. Hawkes
A.C. NIELSEN COMPANY
Nielsen Plaza
Northbrook, IL 60062
USA

Session 16. Evaluating from Household Surveys the Social Consequences of the Structural Adjustments

Mr. Michel Glaude
INSEE
18 Boulevard Adolphe-Pinard
75675 Paris, Cedex 14
France

Session 17. Non response (New Methods and Worldwide Trends)

Dr. Anders Ekholm
UNIVERSITY OF HELSINKI
Department of Statistics
Aleksi 7
SF-00100 Helsinki
Finland

Session 18. Quality Assurance Measures for Surveys

Dr. David A. Binder
STATISTICS CANADA
R.H. Coats Building
11th Floor, Section F
Tunney's Pasture
Ottawa, Ontario
Canada K1A 0T6

Session 19. Effect of Market Deregulation on the Quality of Economic Statistics

Mrs. M. Kollanyi-Juahasz
HUNGARIAN CENTRAL STATISTICAL OFFICE
Keleti Karoly 5-7
Budapest 1024
Hungary

Session 24. Geographic Information Systems in Surveys

Professor Sergio Zani
UNIVERSITA DEGLI STUDI
Instituto di Statistica
Via Kennedy 6
43100 Parma
Italy

Session 30. Surveying and Assessing Marine Populations

Professor K.V.R. Sastry
Annapurna
House No. 8-2-310/B/19
Road 14
Banjara Hills
Hyderabad-500 034 (A.P.)
India

III. FLORENCE SESSION - 1993

3.1 List of IASS Programme Committee members

B.N. CHINNAPPA (Canada, Chair)
K. WOLTER (USA, Chair of 1991 Programme Committee)
O.O. AJAYI (Nigeria)
J.G. BETHLEHEM (Netherlands)
P.P. BIEMER (USA)
W.O. BUSSAB (Brazil)
B.A. CARLSON (UNICEF)
R. CHAMBERS (Australia)

J.C. DEVILLE (France)
T.J. MARCHANT (World Bank)
J. ROY (India)
C.E. SARNDAL (Canada/Sweden)
F.J. SCHEUREN (USA)
C.J. SKINNER (U.K.)
A. THIONGANE (Senegal)
D. WEI (Taïwan-China)
H. WU (China)

3.2 Topics for the 1993 IASS meeting

A list of suggested topics has been sent to all IASS members together with a letter from Mrs B.N. Chinnappa, Chair of the 1993 Programme Committee, seeking comments on these topics. If you have not answered yet, please write directly to Mrs Chinnappa, 9 Goodwin Ave, Nepean K2E 5C8, Canada.

IV. LOCAL REPRESENTATIVES

The Secretariat has been sending to all IASS Local Representatives a package containing a copy of the IASS Statutes, a few application forms, some IASS Stationery, a note about the Association, a document about UNESCO coupons, and several copies of the Survey Statistician ; they can disseminate them in conferences or seminars.

We hope that all these documents will help them in their recruiting task.

V. ANNOUNCEMENTS

5.1 International Conference on Survey Design, Methodology and Analysis, Colchester, 1990.

The IASS is sponsoring an International Conference organized by the Institute of Statisticians on Survey Design, Methodology and Analysis from 4 th to 7 th July 1990 at the University of Essex, Colchester, England.

If you are interested in receiving further information, please write to :

Tessa KONRATH,
Institute of Statisticians
50 Fitzroy Street
London W1P 5HS - England.

5.2 International Conference on Measurement Errors in Surveys

Seeks Contributed Papers

The field of research in measurement errors in surveys spans a number of disciplines including sociology, psychology, statistics, communication science, business, and economics. To document the current state of the field, to report the findings of new research, and to promote interdisciplinary and cross-cultural interchanges, an International Conference on Measurement Errors in Surveys will be held November 11-14, 1990, in Tucson, Arizona, U.S.A., at the Sheraton El conquistador Hotel. The Conference is jointly sponsored by the American Statistical Association (Section on Survey Research Methods), the American Association for Public Opinion Research, the American Marketing Association and the International Association for Survey Statisticians.

The Conference will bring professionals together from different disciplines and from around the world to share approaches to modeling, assessing, and reducing measurement errors in surveys. Approximately 60 invited papers and 125 contributed papers will be presented during the three-day conference. An edited monograph of the best papers will be published as part of the project and sent to each registrant. Information on registration will be published in the newsletters of the four sponsoring organizations.

Abstracts for contributed papers are sought on the following topics :

- the questionnaire as a source of measurement error (e.g., effects of question wording and question order, the role of pilot tests, experiments in questionnaire development),
- the respondent as a source of measurement error (e.g., memory effects, proxy vs. self response, effects on the respondent's mood),
- measurement errors due to data collectors and data collection (e.g., effects of interviewer training and education, reinterviews, evaluation of interviewer performance),
- the respondent-interviewer interaction (e.g., sensitive questions, structured vs. nonstructured interviewing),
- modeling measurement errors and their effects on estimation and data analysis.

Review papers and descriptive papers on methods for dealing with measurement errors may be appropriate if they illustrate widely applicable methods. However, the papers most sought are those containing empirical data.

Send two copies of a 1-2 page abstract to arrive by May 1, 1990, to :

Lee Decker
American Statistical Association
1429 Duke Street
Alexandria, VA 22314, USA

A $100 non-refundable "submission fee" must accompany the abstract (no purchase orders, please). If the abstract is accepted, the $100 will be applied to the $365 conference registration fee. Contributors will be notified of acceptance by June 1. If the abstract is not accepted, the submission fee will be returned.

For more program information, contact Paul P. Biemer, Chair of the Organizing Committee, Experimental Statistics Department, Box 30003, Dept. 3130, New Mexico State University, Las Cruces, NM 88003-0003. For information on registration, contact Lee Decker at the ASA.

5.3 Symposium on Improvement and Measurement of Data Quality

The dates for the previously announced symposium on Improvement and Measurement of Data Quality sponsored and organized by Statistics Canada in Ottawa have now been fixed as October 29 to 31, 1990. Some themes will be : the challenge of improving data quality with reduced resources, measurement of Census coverage and adjustment of estimates, improvement of the quality of frame data, measurement of Census error, improving data collection, issues with use of administrative data and quality control procedures. Invited speakers will be coming from statistical offices in Europe, Australia and North America, from universities, and from the private sector. To receive programme and registration information about this symposium, contact Mary March or Robert Lussier, Methodology Branch, Statistics Canada, Ottawa, Ontario, K1A 0T6. Tel. : 1-613-951-1473.

5.4 Meeting of the American Association for Public Opinion Research

The 1991 meeting of the American Association for Public Opinion Research (AAPOR) will be held May 16-19 in Phoenix, Arizona. The meetings cover methodological advances in data collection and analysis in surveys, and substantive issues on public opinion. The 1992 meeting of AAPOR is to be held jointly with the World Association for Public Opinion Research (WAPOR). IASS members may be interested from the Far and Near East, South America, Canada, and Europe in attendance. AAPOR meets jointly with WAPOR in even numbered years, with the sessions integrated to a certain extent. In 1992, WAPOR's formal meeting will be held on Thursday, May 14, followed by AAPOR's meetings starting the evening of May 14 and running through noon on Sunday, May 17. The 1992 meeting will be held at the Don Cesar Beach Resort in St. Petersburg Beach, FL. Information about meetings are available by writing to AAPOR, P.O. Box 17, Princeton, NJ 08542, USA.

5.5 Conference of the Israeli Statistical Association

The next Annual Conference of the Israeli Statistical Association (ISA) will be held in the Hebrew University in Jerusalem on May 24, 1990.

On the initiative of the ISA, a number of half-day meetings have been held in various institutions where statistical methods are applied, in order to bridge the gap between statistical theory and its application. Recently, two such meetings were held in the Israeli Standards Institute and in the National Metereological Station.

5.6 Meetings of the Committee of Experts in Sociodemographic Statistics

The first and second meetings of the Committee of Experts on Sociodemographic Statistics were held in Buenos Aires in 1989. Documents submitted by national technicians were discussed during these two meetings, with active participation by international experts.

These meetings were held thanks to the financial support of UNFPA (United Nations Fund for Population Activities).

Copies of the documents are still available in both English and Spanish.

Contact :

ENCUESTA PERMANENTE DE HOGARES INDEC
H. IRIGOYEN 250
1310 BUENOS AIRES
ARGENTINA

5.7 Meeting of the Slovak Statistical and Demographic Society

The annual meeting of the Slovak Statistical and Demographic Society was held in Bratislava on March 14, 1990. The new Council has been elected with Mr. Abraham in chair and the programme for the period 1990-94 has been established. For further information, please contact Dr. Miro Abraham, Dean of the Economic Faculty, 974 00 Banska Bystrica, Czechoslovakia.

The first meeting of the Czech Statistical Society was held in Prague March 29th, 1990. This Society is going to co-operate with the Slovak society mentioned above and with other statistical societies in European countries. For further information please contact Prof. Vaclav Cermak, Prague School of Economics, Zapotockeho 4 130 67 Praha 3, Czechoslovakia.

VI. COUNTRY REPORTS

6.1 Philippines

Survey of Household Operated Activities (SHOA, 1988)

The Integrated Survey of Households (ISH), National Statistics Office (NSO), serves as vehicle for the ten-year program on household capability surveys which include labor force and employment, fertility and demography, functional literacy, income and expenditure, household operated activities and others.

The 1988 Survey of Household Operated Activities (SHOA) was conducted in November 1988 with the objective of measuring sectoral contributions of the household operated activities to the total economy. The reference period covered from November 1987 to October 1988. The sampling frame was the Integrated Listing of Households and Establishments conducted in February 1988. This survey used the PSU's (villages) of the ISH except that post stratification sectoral activities of the households within PSU's was introduced. The sampling fraction used for the SHOA was also different from the ISH as the SHOA was to obtain adequate sample households for each sector as much as possible. The sampling design of the SHOA had also provided priority criteria for households with more than one economic activities as follows : Mining and Quarrying, Manufacturing, Wholesale and Retail Trade, Agriculture and Services. There were estimates for each industry and an aggregate estimate for the total contribution of the informal sector to the economy for National Accounts (NA) estimation. SHOA is a new survey for providing data gaps in NA.

6.2 China

I. China's 1990 Population Census

The preparations for 1990 Population Census is well underway. The reference date for the census will be on 1 July 1990. The total population will be more than 1.1 billion.

For details, please contact Mr Shen Yimin, Director of Department of Population Statistics, State Statistical Bureau (SSB), Beijing.

II. New Major Surveys

1. A Pilot Survey on the Food Consumption, Nutrition and Health will be conducted in 1990 in seven provinces, autonomous regions and municipality, namely Beijing, Hebei, Heilongjiang, Zhejiang, Guangdong, Sichuan, Ningxia, with the financial aid from FAO, WHO and UNICEF.

The questionnaire consists of seven parts :

(1) a list of the household members, (2) the state of health of the household members and the information on their going to doctors, hospitalization and payment, (3) smoking, (4) the basic condition of the household, (5) the state of health of the children aged under 6 years, (6) the food items, (7) the purchase, expenditure and consumption of 8 major kinds of food.

35 thousand households with 159 thousand persons in urban and rural areas will be enumerated.

For details, please contact Yang Ye, SSB, Beijing.

2. A survey on the basic situation of children will be conducted in 1992 with the financial aid from UNICEF and G. W. Kalton as the statistical advisor.

For details, please contact Yang Dongni, SSB, Beijing.

6.3 Nigeria

The Design for Nigeria's National Census of Industries and Business (NCIB)

by

J. B. Coker,
Fed. Office Statistics,
Lagos, Nigeria.

1) Objective :

The first National Census of Industries and Businesses (NCIB) is being co-sponsored by Federal Office of Statistics (FOS), the State Statistical Agencies (SSA's) and the Federal Ministry of Industries (FMI).

2) Coverage :

All establishments within the geographical territory of Nigeria are included. These fall into three broad categories : (i) Medium and Large Establishments engaging 10 persons ; (ii) Small Scale Establishments engaging 1-9 persons (SSE) and (iii) Survey of Household enterprises (SHE). The first category is being covered on 100 % basis while the other two are being covered on sample basis.

3) Reference and Survey Period :

1986 was the reference year for the census. Medium and Large Establishments which, almost invariably keep records of their activities, are being covered in 1989 with the questionnaire making information for 1988. The small scale establishments and household enterprises, most of which do not keep records are covered on current basis. Each unit reports for one month, but the units are spread randomly and evenly across the 12 months of the reference year - April, 1988 to March 1989.

4) Sample Design :

The Survey of Small Scale Establishments (SSE) and the Survey of Household Enterprises (SHE) were tagged on to the sample for the National Integrated Survey Households (NISH). The summary of NISH sample design is as follows :

(i) Multi-stage Design - with population census Enumeration Areas (EAs) as the first-stage sampling unit and housing unit as the last stage sampling unit ;

(ii) Two-phase Design - In the first phase, EAs were selected with equal probability while at the second phase, EAs were selected with probability proportional to the measure of size obtained through a quick count of housing units at the first phase ;

(iii) Stratified Sampling - Each state being a domain, separate sample was selected in each. Each state was also divided into 3 sectors based on size of each locality, namely : - urban, semi-urban and rural.

(iv) Replicated and Rotated Design - A sample in 10 independent replicates was chosen for the five year survey period with 6 replicates covered any year. Each year, on of the last years replicates in dropped while another one is chosen from the unused replicates. Each replicate in each state has 8 EAs in the urban, 2 EAs in the semi-urban and 10 EAs in the rural ;

(v) Sample Spread Across 12 Months - The sample EAs were randomly and evenly spread across the twelve months of the year so that each month 10EAs (4 urban, 1 semi-urban and 5 rural) are to be covered in each state ;

(vi) EPSEM Design - Within each sector (urban, semi-urban and rural) in each state, the sampling fractions for selecting housing units were determined so as to achieve a self-weighting sample of housing units.

The sample of housing units in each state was used for the SHE. All enterprises in the sample were covered.

For SSE, however, a listing of establishments was carried out in each EA just before the survey month. The small scale establishments were grouped into two - those engaging less than 5 persons and those engaging 5-9 persons. Different samples were taken in each group. It was not possible to achieve a self-weighting sample of establishments.

5) Estimation Procedure :

With the division of the census universe into 3 non-overlapping components, the separate estimates for the self-weighting design for housing units, the estimates for SHE involves, summing up across all sample EAs in the state and applying a grossing up factor at the state level. For the SSE, however, since the sample of establishments is not self-weighting, weighting has to be done at the EA level before summing up across all sample EAs in the state.

The sample design for the SSE did not use the lists of establishments being compiled because at the start of the NCIB programme, it was not certain whether all states would be able to carry out the listing exercise. Besides, where the listing has been done, the list would be used to improve the estimate for the state.

For the medium and large establishments, since no sampling was involved, the estimation procedure involved only summation across all establishments with adjustments : made for non-response.

6) Method of Data Collection :

For the Medium and Large Establishments, self enumeration is envisaged although field staff would be available to assist respondents in completing the questionnaire where the need arises. NCIB programme also plans for hand delivery of questionnaire to the establishments and hand retrieval too. Although this procedure is expensive but it is expected to appreciably improve the response rate.

For the SHE and SSE, interview method is being used. Since many of the units under study do not keep record, it has been found necessary to provide a daily diary of purchase and receipts to ensure that reliable data are obtained in respect of input costs and gross output. And as was noted in paragraph 3 above, the reference period was reduced to one month for those two categories - (SHE and SSE).

6.4 USA

UNICEF has just completed a global assessment of malnutrition in children under five, based on data available as of 1 January 1990. The analysis is based on nationally representative data from 76 developing countries accounting for 83 % of the under five population of the developing world, excluding China. The data, which come from a variety of sources, refer mainly to national malnutrition prevalences in the 1980s. For more information write to Dr. Beverly Carlson, UNICEF House, Three United Nations Plaza, New York, NY 10017.

UNICEF is preparing guidelines for conducting a school height census. School height censuses can provide information on nutritional status and its causes to assist both decision making and the allocation of government resources to combat malnutrition. School height censuses are increasingly being used as an effective source of information in national food and nutrition surveillance systems. For more information write to Dr Beverly Carlson, UNICEF House, Three United Nations Plaza, New York, NY 10017.

The National Institute on Alcohol Abuse and Alcoholism is sponsoring the National Longitudinal Alcohol Epidemiology Survey. This survey will provide information on the quantity, frequency, and patterns of alcohol use in the U.S. ; the prevalence of alcohol and selected drug use disorders ; the use of alcohol-related treatment services ; and changes in these variables over time. The initial survey of approximately 61,000 housing units will be conducted from September 1990-August 1991. Two follow-up surveys are planned at 2-year intervals. For more information write to Chester Bowie, Chief, Income Surveys Branch, Demographic Surveys Division, U.S. Bureau of the Census , Room 3377-3, Washington, D.C. 20233 U.S.A.

Westat, Inc., is assisting the development of education data collection and reporting standards for use by state, local and federal agencies. The U.S. National Center for education Statistics implemented this project in support of the goal of producing comparable and uniform data on the condition of education in individual states and the nation as a whole. Members from the major users groups were selected to serve on a task force responsible for developing the standards. For more information write to Dr. David Bayless, Project Director, Westat, Inc., 1650 Research Boulevard, Rockville, MD 20850, or call (301) 251-8283.

Recent fellows of the American Statistical Association/National Science Foundation Research Fellowship Program in residence at the U.S. Bureau of the Census and the U.S. Bureau of Labor Statistics include Dr. Malay Ghosh (University of Florida), Dr. Nell Sedransk, (University of Iowa). Dr. Ghosh intends to develop a Bayesian analysis of a general mixed linear model that can be used in developing small area estimates ; Dr. Nell Sedransk will research the use of mixture models as a basis for developing inferential procedures for

Consumer expenditure Survey data, including point and interval estimation of seasonal and non-seasonal trends and tests of hypotheses. Dr. Joseph Sedransk will examine the effects of different imputation methods on confidence intervals for means and regression coefficients when data are not missing at random.

6.5 Israël

A. Sample surveys in the Central Bureau of Statistics

In addition to the regular surveys, a number of other new or periodical surveys have been completed recently, several surveys are at present being conducted and some surveys are now being prepared :

1) Surveys that have been completed :

(a) Survey of Persons Aged 60 and over.
(b) Survey of High-School Teachers of Scientific Subjects.
(c) Survey of Reading and Leisure Activities.
(d) Survey of Reasons for Abstention in the General Elections.

2) Surveys now being conducted :

(a) Survey of Construction Costs.
(b) Survey of Trucks.

3) Surveys in preparation :

(a) Time - Use Survey.
(b) Survey of Absorption of Immigrants.
(c) Use of Health Services Survey.
(d) Victimisation Survey.
(e) Survey of Leisure Activities of High-School Students.

B. Up-dating samples of the following current surveys now being conducted :

1) The set of enterprise surveys :

(a) Survey for the Monthly Labor and Wages Indices.
(b) Survey for Monthly Industrial Indices (productivity, sales,etc.).
(c) Annual Industry and Crafts Survey.

2) Survey of prices for the Consumer Price Index

C. Methodology

1) Redesign of the current Income Surveys : reduction of the reference period from twelve months to three months : revision of imputation methods and development of a new weighting procedure

2) Calculation and presentation of sampling errors : a "friendly" and interactive computer program has been developed for this purpose, to suit the sampling

and estimation methods used in Israel, in a wide spectrum of surveys. The program provides simultaneously all estimates and their sampling errors in a form suitable for publication. The program was used for two recent publications (Income Surveys and the Elderly Survey).

3) Classification of the population into homogeneous groups : In order to improve post-stratification and various imputations, a classification into homogeneous groups was implemented in two surveys (Income Survey and Family Expenditure Survey) on a hierarchical basis ("tree"), instead of the conventional cross-classification.

6.6 Japan

Statistical Study on Japanese National Character
and its International Comparison Among 5 Nations

The statistical study of Japanese national character has continued since 1953. The surveys were done every five years. The last survey was completed in 1988. The sampling procedure as below.

Nation Wide Sample Survey by Stratified Three Stage Random Sampling, Sample Spot : 200-300, Sample Size : 2000-4000,

1st Survey	in 1953,	2nd Survey	in 1958,
3rd Survey	in 1963,	4th Survey	in 1968,
5th Survey	in 1973,	6th Survey	in 1978,
7th Survey	in 1983 and	8th Survey	in 1988.

The data were collected by one-to-one interview method for the sample using a definite questionnaire.

In this study Japanese national character is operationally defined, as belief systems, the way of thinking and emotional attitude of the Japanese.

The questionnaires include three types of items :

(1) those dealing directly with Japanese national character, e.g. actions and/or attitudes which specialists describe as being uniquely Japanese.

(2) items drawn from counterpart foreign survey research.

(3) items designed to yield data that can be compared with foreign studies.

From the point of view of behaviormetrics, national character, defined as above, is represented by data structure and the set of relative frequency distributions of opinion and emotional attitude.

The analysis of time series data reveals various features of the Japanese, including the change and unchange of opinions, their dynamic features by breakdowns, cohort analysis and dynamic analysis of opinion structures.

Thus, the consistent nature of the japanese mind, and signs of formation of new Japanese belief systems and the way of thinking through the statistical analysis of time series data are given. The changes and unchanges are shown as change and unchange of frequency distribution of response and as those of opinion structure represented by data structure.

This study began from the cross-societal research in different cultures, based on the surveys, 1971, Japanese American in Honolulu ; 1978, 1983, Honolulu residents ; 1978, American in mainland ; 1982, French. The new program of this study was realized in 1987 and 1988 under a new idea of cultural link analysis for comparative study (see, The Survey statistician nr 21, 14-15).

This program covers the surveys of Japan, Honolulu residents, America, U.K. France and Germany. The analysis of these data is going on with the past time series data.

For future information, please contact, Chikio Hayashi, The University of the Air, 11, WAKABA-2, CHIBA, 260, JAPAN.

6.7 Canada

Statistics Canada is developing an Address Register for use in improving coverage in the 1991 Census of Population. The Address Register will be a machine readable list of residential addresses in urban areas of Canada. It is being created by merging and unduplicating addresses from administrative data sources including income tax records, municipal assessment rolls, and telephone and hydro company billing files. Its use in the 1991 Census will be to provide an independent source to which the address lists compiled by Census enumerators will be compared and reconciled. Following the 1991 Census, the feasibility of regularly updating the Address Register and using it to replace the current area frame in use for the Labour Force Survey and other household surveys will be studied. For further information on the Address Register, contact Douglas Drew, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6.

The Statistical Society of Canada at its last board meeting, in October 1989, approved the establishments of : i) a Biostatistics Section ; ii) a permanent office. Persons interested in more information about the Society can write to : Statistical Society of Canada, Dunton Tower, 6th Floor, Carleton University, Colonel By Drive, Ottawa, Ontario, K1S 5B6. The comprehensive user-friendly seasonal adjustment software package, X11ARIMA/88, is now available for a wide array of computers, including mainframes and IBM/PC compatible microcomputers. An enriched version of the earlier X11ARIMA, this version features additional ARIMA models for the automatic extrapolation routine, estimation of Easter effects, automatic variable selection of the seasonal moving averages, permanent a priori modifications, automatic removal of trading-day variations and/or Easter effects before ARIMA modelling, new charts and user-specified printouts. For more information on this software, contact Estelle Dagum, Time Series Research and Analysis Division, Statistics Canada, Ottawa, Ontario, K1A 0T6. Tel. : 1-613-951-9876.

6.8 Botswana

Botswana has started preparations for a Population and Housing Census to be conducted in August, 1991. This is in conformity with the intention of Governement to maintain decennial Census which will ensure

the continued availability of demographic and socio-economic bench-mark data at regular intervals in the Government's programme of data collection, processing analyses and dissemination.

The immediate objectives of the 1991 Population and Housing Census are :

(i) to provide demographic, social and economic data to update the 1981 Population and Housing Census data

(ii) to provide current data for the determination of fertility, mortality and migration levels, patterns and trends ; and population growth at both national and sub-national levels

(iii) to interpret and disseminate the population census data in a manner which will ensure their effective applications as well as an under-standing of the population factor in national development programmes.

The Census is being conducted on de facto basis. It is a one-phase census where each household is visited once and there are no additional questions to be asked on a sample basis.

The cartographic/geographic work for the Census started in November, 1989 and it is expected to be completed at least 3 months before the Census date in August 1991.

The questionnaire has already been drafted - it is to be pretested in May, 1990 after which it will be finalised for the Pilot Census in August same year. A Post-Enumeration Coverage Evaluation Survey is being planned for September, 1991 immediately after Census Enumeration.

6.9 Czechoslovakia

Microcensus 1989

In the Czechoslovak official statistics, the microcensuses have been understood as large-scale periodic sample surveys. Their aim is to investigate the basic indicators of the living standard of the Czechoslovak population. Such surveys are made currently in most of the statistically developed countries. In Czechoslovakia, the microcensuses have been made irregularly from 1957 and the 10th round of that survey was realized in March 1989. The momentary indicators, for example some socio-demographic ones, have been related to the December 31, 1988, and the interval indicators, for example the ones about incomes, have been related to the whole year of 1988.

The main purpose of the Microcensus 1989 was the same as for previous ones ; the purpose was to collect statistical data on the social and economic structure of Czechoslovak households in 1988 and on their income differentiation. The sample was created by the method of stratified two-stage sampling and the size of the sample was more than 100,000 households, i.e. 2 per cent of all households in the state.

The results of the Microcensus 1989 are prepared for publication in the Statistical Yearbook of CSSR 1990 and in mimeographed fourth volumes collection (report) of the Czechoslovak Federal Statistical Office.

6.10 Turkey

1) Census of Population 1990

In addition to the recent report (Survey Statistician No.22) on population censuses in Turkey, new developments have taken place. The 1990 Census of Population will be taken on October 21, 1990 on "de facto" basis. There is a tendency towards taking censuses on "de jure" basis when the field conditions are ideal in future rounds.

Inquiries about the census plans can be addressed to Census Division, State Institute of Statistics, Necatibey Caddesi No.114, Ankara, Turkey.

2) Household Labour Force Survey 1988

In Turkey, some information on employment and labour force has been collected periodically by population censuses. Since the beginning of the Five Years Development Plan period, detailed information is required on labour force. The State Institute of Statistics (SIS) has started to carry out the Household Labour Force Survey in 8 industralised cities on pilot basis in 1966 and has started to conduct the surveys periodically since 1967.

In 1988, SIS has started to conduct a new Household Labour Force Survey on national basis twice a year by using a rotation sample design. Report of the 1988 Household Labour Force Survey is now at printing stage.

Further information can be obtained from Labour Statistics Division, State Institute of Statistics, Necatibey Caddesi No.114, Ankara, Turkey.

3) Manpower Survey of Hotel and Tourism Industry 1989

Due to the rapid growth and demands for the tourism industry in Turkey, the Ministry of Tourism has established a Manpower Survey of the Hotel and Tourism Industry in 1989.

The survey covers accomodation establishments, restaurants and travel agencies licenced by the Ministry of Tourism. The sample selection, fieldwork (data collection) and data processing are done by the State Institute of Statistics on behalf of the Ministry of Tourism.

Survey results are published in two volumes (Current Manpower Resources, and Projection of Manpower Requirements) by the Ministry of Tourism.

Further information may be obtained from the Ministry of Tourism, Ankara, Turkey.

4) Monthly Industrial Production Index

State Institute of Statistics has established a new monthly index of Industrial Production, 1986=100 based. The first of the series is released in April 1990 covering the period from January 1986 and February 1990.

The index is based on the monthly production information of 112 main industrial items in three main sectors which cover 60 % of the total industrial production of Turkey.

The index will continue to be published on monthly basis in the future.

Further information can be obtained from the Price Indices Division, State Institute of Statistics, Necatibey Caddesi No.114, Ankara, Turkey.

6.11 Italy

Sample Survey Activity

(a) Edit & Imputation Generalized Programs
(b) Quality Data Control
(c) Small Area Sample on Family Survey (ISTAT)
(d) Composite Estimator AK on Labour Force Survey (ISTAT)
(e) Missing Data : Distortion Effect
(f) Longitudinal Analysis of Labour Force Participation by means of Retrospective Surveys
(g) Variance Estimation of the Consumer Prices Indexes
(h) Imputation Income Estimates : Some Experiments by Multivariate Analysis

Contact, please :

For a) and b) points : M. Masselli, National Statistical Bureau, Via Cesare Balbo, 16, 00100 Rome, Italy (Tel. +/6/4673) ;

For c), d) and e) points : A. Russo, National Statistical Bureau, Via Cesare Balbo 16, 00100 Rome, Italy ;

For g) points : U. Trivellato, Department of Statistics, Via S. Francesco 33, 35121 Padua, Italy (Tel. +/49/657531) ;

For g) point : L. Biggeri, Department of Statistics, Via Curtatone 1, 50123 Florence, Italy (Tel. +/55/216431) ;

For h) point : D. Lucev, Institute of Statistics, I.U.N., Via Acton 38, 80133 Naples, Italy (Tel. +/81/5522588).

6.12 German Federal Republic

A meeting of the section of the German Statistical Society "Methodik statistischer Erhebungen" took place in Duisburg on 22 September 1989 with H. Strecker in the chair and with the following 4 papers read :

1) H. Schneeberger, Universität Erlangen-Nürnberg : Optimierung in der Stichprobentheorie (Optimizing in Sampling Theory) - It was shown that in optimum allocation with $n_h > N_h$ the method to take $n_h = N_h$

yields the same result as using the gradient-projection method, and that this optimum is global and feasible. "Optimum" stratification according to Dalenius' formula can yield (local and global) minima, maxima, saddle-points and non-feasible solutions at all.

2) J. Schmidt, Statistisches Bundesamt, Wiesbaden : Der neue Stichprobenplan des Mikrozensus ab 1990 (The New Sample Design of the microcensus as from 1990) - The most relevant items are described in the report "New sample designs for 1990".

3) W. Foerster, Universität Marburg : Aubenstand des Leergutes einer Brauerei - Ermittlung über eine Stichprobe der Rücklaufzeiten (The Evaluation of Outstanding Claims of Breweries' Empties by a Sample Estimate of the Average circulation Period of Bottles) - It is assumed that circulation of single bottles can be described by an unique probability model. A two stage sampling procedure including regional stratification has proved to be successful in a case study.

4) A. Diekmann, Universität Mannheim : Probleme in der Umfrageforschung (Problems in Empirical Research) - e.g. statistical problems with retrospectively assessed data in surveys and panel studies, computer supported techniques in surveys conducted by telephone.

Report on New Sample Designs for 1990

In 1990, methodological interest in relation to sample surveys of official statistics is centred on the microcensus and the survey of salary and wage structure.

The microcensus as the largest population sample survey taken on a regular basis provides every year socio-economic structural data on the population and its changes ; it helps to currently observe the labour market and it is also used as a grossing-up frame for numerous samples in empirical social research.

A new sample design has been developed for the microcensuses to be taken as of 1990 which is to replace the former one used since 1972. With a constant sampling fraction of 1 % per year, the new sample design will permit above all to achieve a more detailed regional analysis of the results down to below the level of administrative districts (Regierungsbezirke). The microcensus will continue to be taken as a one-stage cluster sample, but in view of the survey objectives the size of the clusters has been substantially reduced from an average of some 23 dwellings to about 9 dwellings. The sample districts (clusters) are composed of entire buildings or parts of buildings. To this end, the buildings of the sampling frame are first subdivided into four strata. Three strata are classified according to the number of dwellings per building, while one stratum is set aside for buildings comprising collective housing units. This material stratification is integrated into a detailed regional stratification including spatial units with an average of

about 250,000 inhabitants. The sampling frame is based on up-to-date figures on the number of dwellings per street and house number which are available from the 1987 Population Census.

For the first time since 1978, a survey of the salary and wage structure will be taken in 1990. The survey provides information on the structure and distribution of wages and salaries of wage earners and salaried employees in production industries, commerce and the credit and insurance business.

The survey is conducted as two-stage sample. At the first stage establishments are selected, at the second stage employees. Pursuant to the relevant legal provisions, a maximum of 600,000 employees may be included in the sample. Prior to the sampling, the establishments included in the sampling frame are stratified for each federal Land by economic sectors and for each sector by 6 size classes of employees. The allocation of the overall sample size to the individual strata follows the preset precision grades for the estimated gross monthly earnings of the employees. The sampling frame is provided by the 1987 Census of Non-Agricultural Local Units and the file of establishments in production industries.

For more detailed information on the sampling methods used for the above surveys please contact Mr Jürgen Schmidt, Subdivision Mathematical-Statistical Methods, Federal Statistical Office, D 6200 Wiesbaden, Federal Republic of Germany.

6.13 Hungary

Sample Surveys

In our days no report can be written on Hungary without reference to the odd changes in politics, economy and society which are underway in Eastern Europe and make headlines in the press all over the world. This sketch is therefore split into two parts : the first is devoted to what has happened before the changes, and the second to what is done now and might be expected in the future.

Recent Past : sample surveys were relatively rare, and the most important ones were conducted by the CSO. Of the few two are mentioned here : (i) the Unified System of Household Surveys (USHS), and (ii) livestock survey. The main features of (i) are as follows :

- a multi-purpose .3 - .4 p.c. rotating sample of the population based on the decennial censuses ;
- a carefully organized ten-year-schedule of household budget-, labour force- and socio-statistical surveys ;
- two- and three-stage sample design, two-way stratification of the sample ;
- a permanent staff of subject-matter statisticians and a permanent network of 430 interviewers responsible for devising, coordinating and conducting the surveys.

(ii) works on a 4 p.c. sample of all non-state-owned and non-cooperative-owned farms. The two-stage two-strata sample was based on tri-operationnal censuses and, using a linear regression estimator, yielded

estimates for the main indicators (pig and beef) with a coefficient of variation of about 1 p.c.

Present and Future : Sample surveys will have an increasing role, they should replace censuses in several fields of economic statistics. In the USHS the sample design will be updated and improved to meet the increased demand on household and labour force surveys ; the latter should be made a reliable tool of observing labour market and unemployment. A lot of modern and streamlined techniques should be adopted to handle the increased volume of tasks at a time when the willingness of population to cooperate with the interviewers will certainly deteriorate.

6.14 Syria

The Central Bureau of Statistics works on carrying out specialized field surveys within specific periods which conducted according to time schedule and in an agreement with the allocated financial credit.

The Central Bureau of Statistics has conducted the following Statistical surveys :

1) Survey of planted areas of wheat, barely, lentil and chiek. Peas . Annual sample survey for estimating planted areas, production and fields of crops. This survey is carried out in cooperation with the Ministry of Agriculture and Agrarian Reform and the General Union of peasants.

2) Industrial survey of private sector Establishments and Institutions which is conducted according to selected samples. Computer will help in completing framework.

3) Index number of wholesale and retail prices survey with reconsidering calculation procedures relating to wholesale and retail prices.

4) Solar Energy survey.

5) Index number survey of cost of residential building per square metres

6) Survey of average wages and actual work hours of building and construction sector.

7) Man power survey 1989 which aims at measuring size of human power, force power besides employment and unemployment, internal migration.

6.15 Spain

Report of the Basque Country of Spain, by Anjeles Iztueta

EUSKERA

The own language of the Basque Country is the Euskera, spoken currently only by the 25 % of the population. This millennial no Indo-European language is at present in a process of retrieval of its current minorized social position. In this normalizing effort, the school plays an active role, with three types of bilingual programmes in rule at the moment (A,B,D). The study of the evolution, affecting factors, results, etc. of these programmes of bilingual education is already in its third stage, called EIFE-3, in which elaborated sampling technics, multi-variant data analyse etc. have been used.

For more information, please write to M. J. SIERRA. Dpto. de Educación, Universidades e Investigación. Duque de Wellington, 2. 01011 VITORIA-GASTEIZ. Tel. : 45-24.60.00.

OFFICIAL STATISTICS INTERNATIONAL MEETING

In September of 1989 took place in Bilbao an International Meeting of Official Statistics, on three Subjects : Statistical methodology of the Regional Accounts, Experiences in the process of automation of the statistical production, Social Welfare statistical indicators attainment.

The meeting, organized by EUSTAT and IAOS, was attended by 150 specialists from 21 countries all over the world and the minutes are available at : EUSTAT - Instituto Vasco de Estadistica. Dato, 14-16, 01005 VITORIA-GASTEIZ. Spain.

METHODOLOGY PUBLICATIONS

- Statistical operations for sampling. L. Kish. International Statistics Seminar 1986.

- Les enquêtes téléphoniques. V. Salvy. International Statistics Seminar 1988.

- Sampling and non-sampling errors in surveys. A. Marton. International Statistics Seminar 1988.

- New technologies in computer assisted survey processing. Keller and Bethlehem. International Statistics Seminar 1989.

- Generalized Linear Models in epidemiology. J.C. Duffy. International Statistics Seminar 1989.

If interested, they can be acquired at : EUSTAT. Dato, 14-16. 1005 VITORIA-GASTEIZ. Spain.

6.16 Argentina

ON GOING HOUSEHOLD SURVEY EVALUATION PROCESS

The National Directorate of Sociodemographic Statistics (DNES) of the National Institute of Statistics and Censuses (INDEC) initiated an evaluation group that began its activities in October 1987. The main objective pursued by DNES is to establish a planning, control and on going evaluation system to ensure that data is of higher quality and better meets the needs of users.

Activities have begun in the Ongoing Household Survey Department and are designed to evaluate the data organization and collection process, seeking to identify formal, operational, technical and organizational problems that prevent the correct production of information in order to guide the course of action adopted in future survey efforts.

The methodological concept combines qualitative and quantitative techniques, carefully defining specific work areas and establishing strict rules for the recording and systematic analysis of information.

The main techniques used are :

a) observation
b) various types of interviews

c) reinterview survey
d) "weighted collective judgment."

The evaluation strategy defined involves careful study of activities before, during and after the collection of information.

This process brings together producers themselves at all stages (surveyors, supervisors, professional technicians, administrators and executives), promoting closer ties between field personnel and the various departments associated with the survey.

The conclusions and recommendations of this study have proven beneficial to the changes made in field and clerical work (staff training and retraining, reorganization of questionnaire distribution and retrieval, supervision, development of performance standards for enumerators, direct entry of questionnaires on microcomputers) and in relations between the department organizing the survey and the other units concerned.

The OHS* evaluation process has been implemented in the province of Cordoba since May 1989, and trained staff is now available in six other provinces, which will be able to begin their activities in May 1990.

Fore more information, contact :

ENCUESTA PERMANENTE DE HOGARES INDEC
H. IRIGOYEN 250
1310 BUENOS AIRES
ARGENTINA

1990 CENSUS DATA QUALITY EVALUATION PROGRAM

The National Directorate of Statistical Studies of the National Institute of Statistics and Censuses (INDEC) has created an evaluation group which began its activities in July 1988. Two general objectives were identified : the first, immediate objective was to study the efficiency and effectiveness of specific procedures designed to improve the quality of data during census operations, measured in terms of errors not due to sampling ; the second, long-term objective for future censuses was to estimate errors not due to sampling that affect census data and to alert the user to the source and significance of these errors.

To achieve the first objective, three quantitative studies corresponding to each of the pilot censuses were done using reinterview surveys :

1 - To estimate response error for the variables examined in the pilot census.

2 - To identify the sources of errors, that is, to determine the behaviour of specific groups of individuals or variables and to be able to minimize these sources of errors accordingly.

3 - To provide information that generally contributes to the improvement of the census program.

Subsequent phases, making use of exploratory data analysis techniques, examined assumptions about the behaviour of

---

* Ongoing Household Survey

enumerators in the course of their visits and the relationship between control of their training and their performance, and carried out a study that made it possible to classify them based on their behaviour.

The conclusions drawn from these activities led to changes in the questionnaire (making the concepts and definitions used more operational), the layout of the questionnaire, the definition of the duties of enumerators and their training. They also permitted the study of adequate motivation among those surveyed.

As for the second objective, provision was made to conduct a post-census survey to estimate :

1 - Error of omission, which is believed to account for the largest proportion of total coverage error (de facto census).

2 - Response error for some of the variables observed in the census. Estimates of these errors are planned for cities with more than 1,000,000 inhabitants and for the capitals of provinces combined into regions, with distribution by sex and age grouping.

For more information, contact :

DIRECCION NACIONAL DE ESTUDIOS ESTADISTICOS INDEC
H. IRIGOYEN 250
PISO 12 - OFICINA 1225
1310 BUENOS AIRES
ARGENTINA

VII. STATISTICAL ABSTRACTS

7.1 Combination of Surveys and Registers : a Calibration Approach with Categorical Variables

Authors : Johan Heldal[1] and Emil Spjøtvoll[2]

Summary

Inaccurate registers combined with surveys may be used to obtain improved estimates of finite population totals. A measurement error model for the data in the register is introduced. It is shown that the situation can be viewed as a calibration problem for discrete variables. Maximum likelihood estimators, analogues to classical calibration and inverse calibration estimators are studied, both theoretically and by Monte Carlo simulations. For the cases studied, the inverse calibration estimator is as good as the maximum likelihood estimators.

7.2 Record Linkage, Privacy and Statistical Policy

Authors : Lawrence H. Cox and Robert F. Boruch

Journal of Official Statistics, Vol 4, N° 1, 1988 pp3-16, Statistics Sweden.

Abstract

Record linkage technology is a valuable statistical tool. It can be used to enhance the quality, completeness and usefulness of data.

---

[1] Central Bureau of Statistics, P.B. 8131, Dep., N-0033 Oslo 1, Norway.

[2] Norwegian Institute of Technology, N-7034 Trondheim NTH, Norway.

Record linkage also can be used to check the accuracy of data, demanding fewer resources and imposing less respondent burden than other verification methods. In developing formal record linkage policy or informal policy guidelines, organizations must cope with several competing issues regarding the need for privacy and the benefits and costs of record linkage. This paper discusses recent approaches to using the technology and addressing the problems it generates. Our observations are presented from the perspectives of both data user and data supplier.

Key words : Record linkage : statistical matching : exact matching.

7.3 An Evaluation of Statistical Software Procedures Appropriate for the Regression Analysis of Complex Survey Data

Authors : Steven B. Cohen, Judy A. Xanthopoulos, and Gretchen K. Jones .

Journal of Official Statistics, Vol 4, n° 1, 1988, pp17-34, Statistics Sweden.

Abstracts

Data from complex survey designs require special consideration with regard to variance estimation and analysis, because of design components that include unequal selection probabilities, stratification, and clustering. Statistical software package programs are currently available which accommodate a complex survey design, and allow for the generation of centrality parameters and variance estimates for statistics expressed in terms of means, totals, ratios, and multivariate regression coefficients. The methods of variance estimation include the Taylor series linearization method and balanced repeated replication. Using data from the National Medical Care Expenditure Survey, which is characterized by a highly complex survey design, the following four statistical programs appropriate for multivariate analysis of complex survey data are compared : SURREGR, SUPERCARP, REPERR, and NASSREG. The comparison focuses on cost-efficiency, user facility, and program capabilities for a series of regression analyses that are representative of the analytical requirements of the National Medical Care Expenditure Survey.

Key words : SURREGR, SUPERCARP, REPERR, NASSREG.

7.4 On estimating census undercount in small areas

Authors : Cary T. Isaki, Linda K. Schultz, Gregg J. Diffendal, and Elisabeth T. Huang

Journal of Official Statistics, Vol 4, n° 2, 1988, pp.95-112, Statistics Sweden

Abstract

Net undercount rates in the U.S. decennial census have been steadily declining over the last several censuses. Differential undercounts among race groups and geographic areas, however, appear to persist. In the following, we examine and compare several methodologies for providing small area estimates of census coverage by constructing artificial populations. Measures of performance are also introduced to assess the various small area estimates. Synthetic estimation in combination with regression modelling provide the best results over the methods considered. Sampling error effects are also simulated. The results form the basis for determining coverage evaluation survey small area estimates of the 1990 decennial census.

Key words : Census : undercount : adjustment : small area estimation : synthetic estimation : regression : artificial population : simulation.

7.5 On autoregressive model identification

Author : Ette Harrison Etuk

Journal of Official Statistics, Vol 4, n° 2, pp.113-124, Statistics Sweden.

Abstract

Since Cleveland (1972) introduced the inverse autocorrelation function, it has been recognized as a competitor to the partial autocorrelation function as a time series model identification tool. By using simulated and real data, we have demonstrated that neither of these is consistently more powerful than the other for identification of autoregressive (AR) models. However when the underlying AT process is of full order, the partial autocorrelation function invariably is the superior. But when a subset order AR model generates the data, the inverse autocorrelation function is generally more informative. On the whole the partial autocorrelation function exhibits better performance. For instance, in two of the three cases of real series used it clearly outperforms the inverse autorrelation function.

Key words : Autoregressive model identification : partial autocorrelation function ; inverse autocorrelation function.

7.6 The Seasonal Adjustment Procedures for the Consumer Price Indexes : Some Empirical Results

Author : Raj K. Jain, Division of Price and Index Number Research, U.S. Bureau of Labor Statistics, Washington DC 20212,USA.

Journal of Business & Economic Statistics, October 1989, Vol 7(4), pp.461-469.

Beginning with January of 1987, the consumer price indexes (CPI's) have been seasonally adjusted by the X-11 autoregressive integrated moving average (ARIMA) procedure. This modification of the X-11 procedure was introduced following an empirical investigation into three aspects of seasonal adjustment methodology as applied to several CPI series - the choice of ARIMA models to fit and forecast those series, the improvements made by the ARIMA modification in terms of revision and smoothness of the seasonally adjusted series, and the effect on the quality of seasonal adjustment and the identifiability of seasonality due to the ARIMA modification. This article reports the results on that investigation. In addition, a brief account is given of the U.S. Bureau of Labor Statistics procedures relating to the projected seasonal factors, seasonally adjusted aggregate series, and the revisions of the seasonally adjusted series.

7.7 Reduction of Nonresponse Bias Through Regression Estimation

Author : Jelke G. Bethlehem

Journal of Official Statistics, Vol 4, n° 3, 1988, pp.251-260, Statistics Sweden.

Abstract

To investigate the properties of estimators of population characteristics, nonresponse is incorporated in the sampling theory by the introduction of response probabilities. Within this framework the characteristics of the Horvitz-Thompson estimator, generalized regression estimator and post-stratification estimator are studied. It is shown that proper use of auxiliary information can reduce the nonresponse bias.

Key words : Nonresponse ; regression estimator ; post-stratification.

7.8 Computer Selection of Size-Biased Samples

Author : David R. Fox

Journal : The American Statistician, August 1989, Vol 43, n° 3, pp.168-171.

This article describes a method of producing size-biased probability samples as originally proposed by Hanurav (1967) and Vijayan (1968). The aim is to obtain a sample drawn without replacement such that the probability that the i^{th} population unit is included in the sample is proportional to x_i. The complexity of the procedure has led to the development of microcomputer software that greatly facilitates the production of sampling plans as well as the computation of population estimates.

7.9 How Appropriate are Popular Sample Size Formulas ?

Authors : Lawrence L. Kupper and Kerry B. Hafner

Journal : The American Statistician, May 1989, Vol 43, n° 2, pp.101-105.

One concern in the early stages of study planning and design is the minimum sample size needed to provide statistically credible results. This minimum sample size is usually determined via the use of simple formulas or, equivalently, from tables. The more popular formulas, however, involve largesample approximations and hence may underestimate required sample sizes. This article provides empirical evidence indicating that this underestimation phenomenon is extreme for certain sample size formulas based on confidence interval width. Common sample size formulas that consider statistical power are also discussed ; these are shown to perform quite well, even for small sample size situations.

7.10 On Chaos's unequal probability sampling plan

Author : S. Sengupta, Department of Statistics, Calcutta University, Calcutta 700 019, India

Journal : Biometrika (1989), 76, 1,pp.192-196.

Summary

We give necessary and sufficient conditions for the second-order inclusion probabilities to be strictly positive under Chao's (1982) sampling plan for some order of arrangement of the population units. We also prove that the variance of the Horvitz-Thompson estimator of the population total based on this sampling plan is smaller than the variance of the Hansen-Hurwitz estimator based on a probability proportional to size with replacement sample involving the same number of draws.

7.11 The weighted residual technique for estimating the variance of the general regression estimator of the finite population total

Authors : Carl-Erik Särndal, Département de mathématiques et de statistique, Université de Montréal, Montréal, Québec H3C 3J7, Canada ; Bengt Swensson, Institutionen för dataanalys, Högskolan i Orebro, S-701 30 Orebro, Sweden and Jan H. Wretman, U/STM, Statistics Sweden, S-115 81 Stockholm, Sweden.

Journal : Biometrika (1989), 76, 3, pp.527-37

Summary

The paper deals with design based estimation of the variance of the general regression estimator of the finite population total. The usual Taylor linearization variance estimator is an expression in the design weighted regression residuals, in many applications the resulting expression is counterintuitive from a model based standpoint. The improved variance estimator in this paper attaches another simple weight, called "g-weight", to each individual residual. This new variance estimator (i) gives valid design-based confidence intervals, (ii) is nearly unbiased under a suitably chosen regression model, and (iii) works well for conditional inference. Examples are given.

7.12 Assessing the Quality of Household Panel Data : The Case of the Panel Study of Income Dynamics

Authors : Greg J. Duncan, Survey Research Center, University of Michigan, Ann Arbor, MI 48106, USA ; Daniel H. Hill, Opinion Research Institute, University of Toledo, Toledo, OH 43606, USA.

Journal of Business & Economic Statistics, October 1989, Vol 7(4), pp.441-452.

Evidence from a number of methodological studies are used to assess the overall quality of data from the Panel Study of Income Dynamics (PSID). Despite substantial cumulative nonresponse over the nearly two decades spanned by the study, the sample is found to maintain its representation of the nonimmigrant population of the United States. The most important reasons for this result are that the study's following rules insure that the sample replaces itself in the same manner as the population (through the formation of new families by the offspring of old) and that nonresponse is largely unsystematic. Nonresponse also appears to be largely random with respect to parameters in a number of behavioral models. The accuracy of measures is assessed by comparing survey measures with national aggregates and with highly accurate individual validating data. PSID reports of transfer income appear to compare more favorably with program aggregates than do reports from other large-scale surveys such as the Current Population Survey of the United States. Finally, although PSID survey measures generally are unbiased when compared to validating data, they contain amounts of measurement-error variance that range from trivially small to very large.

VIII. QUESTION/ANSWER

Conducted by Leslie Kish. Please send Questions to him (ISR - The University of Michigan, Ann Arbor, MI 48106, USA, TELEX 4320815, FAX 313-747-4575) or to IASS, Paris. Please indicate whether or not you want your name given with the question. This has become an open forum, and we shall gladly print (after refereeing) additions, modifications, discussions of past published answers. Contributors to answers will be acknowledged if they agree.

23.1 Question : The usual formula for computing variances from n observations calls for dividing $\sum (y_i - \bar{y})^2$ by (n - 1). Why is that ? Dividing by n seems more natural and that approach also seems to be supported by some statisticians. Some simple guidance would be appreciated.

Answer. Whether to use

$$s^2 = \sum (y_j - \bar{y})^2/(n - 1) \text{ or } \tilde{s}^2 = \sum (y_j - \bar{y})^2/n$$

is still debated off and on in statistical books and articles, and always assuming I.I.D. random variables, which translates to simple random sampling (srs) in survey terms. With more than 30 (say) independent observations the factor (n - 1)/n may not be important, so we may leave that subject, also the factor (1 - f), for technical treatments. Arguments based on (maximum) likelihood favor $\tilde{s}^2$ and n. But s^2 and (n - 1) are needed for unbiased estimators of the population values S^2 and σ^2, and this is shown for srs, from finite populations in most textbooks on survey sampling (e.g. kish 1965, 2.8.).

However, for the complex samples of survey sampling other problems arise, because the n selections are not independent, not srs. The common use of "paired selections", with two replicates per stratum, provide vivid illustrations : the ratio (n - 1)/n = 1/2 within each stratum denotes the bias in variance estimates if n were used. In survey sampling (n - 1) is used for unbiased estimates.

Two other problems are much more important in complex survey sampling : "design effects" from complex and clustered selections and weighted samples. Consider first "self weighted" samples, based on n selections with equal probabilities (EPSEM). We have these expectations regardless of the complexity of the sample

$$\text{Exp}(\tilde{s}^2) = \text{Exp}[\sum y_j^2/n - \bar{y}^2] = \sum Y_i^2/N - \text{Exp}[\bar{y}^2]$$
$$= [\sum Y_i^2/N - \bar{Y}^2] - \text{Exp}(\bar{y}^2) + \bar{Y}^2 = S^2 - \text{Var}(\bar{y})$$

This justification relies on $\text{Exp}(\bar{y}) = \bar{Y}$, $\text{Exp}(\sum y_j^2/n) = \sum Y_i^2/N$, and $\text{Var}(\bar{y}) = \text{Exp}(\bar{y}^2) - \bar{Y}^2$. It is also conditional on a fixed sample size n, and neglects the factor (N - 1)/N for total population size. Now assume that approximately

$\text{Exp}[\text{var}(\bar{y})] = \text{Var}(\bar{y})$, and consider complex samples in terms of the "design effect" = $\text{deft}^2 = \text{Var}(\bar{y})/(S^2/n)$.

Then $\text{Exp}(\tilde{s}^2) = S^2 - \text{deft}^2 S^2/n = S^2(1 - \text{deft}^2/n)$.

Thus, in complex samples with epsem selections we need merely compute $\tilde{s}^2$ and either $\text{var}(\bar{y})$ or $\text{deft}^2 S^2/n$ to get a good estimate of the element variance S^2. In srs the $\text{deft}^2 = 1$, and we get $\text{Exp}(\tilde{s}^2) = S^2(1 - 1/n)$ and $\text{Exp}(s^2) = S^2$, as we would. Generally the term $(1 - \text{deft}^2/n)$ is unimportant. But it most important that the $\tilde{s}^2$ can be computed simply and used in the denominator of deft^2, regardless of the complexity of the epsem selection.

For weighted estimates the estimate of the variance may be computed as

$$\tilde{s}_w^2 = [\sum w_j \sum w_j y_j^2 - (\sum w_j y_j)^2]/(\sum w_j)^2.$$

Computing "effective n" is less important and more complicated for this brief comment.

23.2 Question : In our work in biological and agricultural applications, when we calculate the "desired sample size" we

obtain huge sizes that we cannot afford. Sometimes the data are skewed or have a negative exponential distribution ; and high coefficients of variation. Can we transform our data (perhaps logarithmically) to get a more normal distribution with lower CV ? The normality would also help with the central limit approach we so much need in survey sampling.

Answer. Yes, highly skewed distributions of quantitative variables are common in statistical and survey data. Not only are income and wealth highly skewed, but also days spent in hospitals or number of books read in a year also. Furthermore, the sizes and activities of firms, organizations, institutions and most economic and social units are highly skewed. There exists a large literature on transformations, also on outliers, which should be related in practical applications. Those articles and books are scattered over the many fields of statistics, but not in the survey sampling literature ; and they are based on I.I.D. assumptions.

Survey sampling, however, is more concerned with complex selections, usually stratified. Also it is increasingly multipurpose in practice. Stratification and "optimal" allocation are the alternative methods for dealing with skewed distributions that we find in survey sampling. But we should note that stratification requires knowledge of some strongly related auxiliary variable(s) in the whole population before selection ; and poststratification needs it after selection. Transformations could work without that population knowledge. An article that shows the possible reductions in the needed sample sizes, due to the reduced variance of C.V. in the transformed variable(s), should be welcome.

I omit from this brief discussion four general obstacles, objections, problems to the "solution" proposed.

a) There maybe substantive obstacles to be acceptance of the transformed variable(s). These obstacles arise in general to all transformations, and I can add nothing here to the general discussion.

b) "The required sample size" appears in all textbooks, something in the form $n = S^2/R^2$ or $n = C^2/CV^2$, where R or CV denoted some "required" precision stated as standard errors or coefficient of variation. The results of these expressions is usually a "required sample size" that is much bigger (by 10^2 or 10^3) than the budget will allow. Then the R an CV are reduced by a factor of ten and the resulting "required" n by 100. This has been my experience and those of my colleagues. Why then go through the exercise ? Additional problems involve (a) a "design effect in D^2S^2 for S^2, also good guesses about S^2, and at the last the "finite population correction" (1 - f).

c) In complex samples "sample size" involves much more than merely the number of elements, n. In multistage samples we must consider the number of units at every stage, but especially the PSU's at the primary stages. And there are many other factors you can find in sampling textbooks.

d) Most surveys are or should be multipurpose so that required samples sizes should have flexibility as to variables, statistics, and domains.

e) The Central Limit question is interesting but must be postponed.

IX. PUBLICATIONS

9.1 Special Issue of the Canadian Journal of Statistics, 1988, Vol 16, Supplement.

The Special Issue of the Canadian Journal was first proposed in the winter of 1987, and a panel of senior professional staff at Statistics Canada was established to serve as a guest editorial board. The objective in producing such a special issue was to display to the professional statistical community a cross-section of work being done within Canada's national agency, and thus enhance the flow of information between academic and operational statisticians.

The result is a special issue which is distinctive in several respects. Its entire complement of articles has been written by statisticians and social scientists associated with Statistics Canada. It contains a paper jointly written by the present and the past Chief Statistician of Canada. The issue contains fourteen papers, whose diversity is extremely wide. As noted from the table of contents, they span time-series analysis, application of the bootstrap method, the use of a new technology such as remote sensing and computer-assisted mapping for statistical purposes, and on-going concerns of applied statisticians such as data editing and imputation and the treatment of survey error, to mention only some of the topics.

The issue may be purchased for $25 (Canadian) from

The Canadian Journal of Statistics
c\o David F. Bray
679 Denbury Avenue
Ottawa, Ontario, Canada K2A 2P2

9.2 Joint Centre for Survey Methods Newsletter, Vol 10, n° 1

Report of the ESRC Seminar, October 1989
Issues in Researching Sexual Behaviour
Social and Community Planning Research,
35 Northampton Square, London EC1V OAX, United Kingdom

9.3 UNESCO's IDAMS Software

IDAMS is a UNESCO software for IBM compatible mainframe computers and microcomputers.

For information contact :
Mme Nicole VISART, UNESCO, IPS/SDA, Place de Fontenoy,
Paris 75700, France Tel : (1) 45 68 38 97/96