INTERNATIONAL STATISTICAL INSTITUTE

Nº 24
DECEMBER 1990

THE SURVEY STATISTICIAN

INTERNATIONAL ASSOCIATION
OF SURVEY STATISTICIANS

Editor : Anne Marie Vespa-Leyder

Association Internationale
des Statisticiens d'Enquêtes
A.I.S.E. – I.A.S.S
c/o INSEE – 18, Bd Adolphe Pinard
75675 PARIS CEDEX 14

C O N T E N T S

****

I. EDITORIAL

As in the past, this issue contains concise and up-to-date country reports which describe various survey and census experiences and make each other aware of what is practised around the world.

There is also an interesting feature on survey sampling problems entitled "Question/Answer" by Prof. L. Kish.

As to presentation, an effort has been made to make the journal's cover more attractive.

Gildas Roy, Editor of The Survey Statistician for several years, has had to give up this post for professional reasons and it is with pleasure that I take over from him. I would like to thank the President, B. Bailar, and B. Grais for this.

The editorial staff of the Survey Statistician offers all IASS members best wishes for 1991.

Editor, The Survey Statistician
Anne Marie VESPA-LEYDER

II. NEWS FROM THE ASSOCIATION/GENERAL INFORMATION

2.1 Dues

A call for dues is being sent to all members for the year 1991. The amount remains F.F. 130 (or F.F. 65 for members from ddeveloping countries).

The members can subscribe to "Survey Methodology", a publication of Statistics Canada at the reduced rate of F.F. 150, or F.F. 75 for members from developing countries (the Association being charged for the difference). We apologize for the mistake in the dues call letter for 1991 mentioning the former rate. (See announcement on cover page 4)

2.2 About the Survey Statistician

Mr. Gildas ROY has had, for professional reasons, to give up the edition of the Survey Statistician. He has been replaced by Anne Marie VESPA-LEYDER as editor of the journal.

In issue n° 23, D. HOLT asked the members for comments about the presentation and contents of our journal. Please go on sending him suggestions so that the publication can be improved.

INTERNATIONAL STATISTICAL INSTITUTE

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

Two Day Workshop on
SURVEY SAMPLING IN DEVELOPING COUNTRIES
Cairo, September 7 - 8, 1991

Presented by: Graham Kalton, University of Michigan
Colm O'Muircheartaigh, London School of Economics
Nanjamma Chinnappa, Statistics Canada
Anis Maitra, United Nations Statistical Office

This workshop will be held just prior to the 48th Session of the International Statistical Institute. 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.

Participants should have the ability to work in English.

Topics to be covered will include:

- frame construction
- sampling methods, including stratification, multi-stage sampling and area sampling
- maintaining control of sampling processes
- sampling weights and variance estimation
- evaluation methods

The registration fee will be $300.00 (US).

If you would like more information about this Workshop, please complete and return the slip below. Please complete promptly as space will be limited.

PLEASE RETURN THIS SLIP TO:

David A. Binder
Scientific Secretary, IASS
11-F R.H Coats Building
Statistics Canada
Ottawa, Ontario, CANADA
K1A 0T6

FAX: 1-613-951-1462

I am interested in receiving further information about the IASS Workshop on SURVEY SAMPLING IN DEVELOPING COUNTRIES, September 7 - 8, 1991 to be held in Cairo, Egypt.

Name _____
Address _____
_____
_____
_____

2.3 Erratum

IASS Members have received the booklets of invited papers from the ISI in Paris (1989). In volume II, Dr. G. Kalton noticed that the pages are out of order. The correct order is : 437, 440, 441, 439, 438, 442, etc. We apologize for that mistake.

III. CAIRO SESSION-1991

The 48th Session of the International Statistical Institute will be held in Cairo (Egypt) from September 9 to 17, 1991. The members will receive Bulletin n° 1 for registration to the Conference. A two day workshop on survey sampling in developing countries will take place just prior to the ISI meeting on September 7 and 8, 1991 (see notice on opposite page).

IV. LOCAL REPRESENTATIVES

The IASS has now a Local Representative in Algeria : Mr. Elyes BENTABET.

The main task of a Representative is to promote the Association in his country, to recruit members and to maintain links between members and the Secretariat and between members themselves. The Association would like to develop its net of Representatives ; therefore members from countries where there is no Representative are invited to contact the IASS Secretariat and to propose themselves for this task.

V. ANNOUNCEMENTS

5.1 Tenth Inter-American Statistical Conference-Aguascalientes, Mexico

The Tenth Inter-American Statistical Session (a specialized conference of the organization of American States) will take place in Aguascalientes (Mexico) in the Headquarters building of the INEGI (Instituto Nacional de Estadística, Geografía e Informatica) from November 13 to 16, 1990.

The National Coordinator is Lic. Mario Palma Rojo, Executive Coordinator of the INEGI, Av. Héroe de Nacozari 2301, Aguascalientes, AGS. Members who are interested in participating should write as soon as possible to the General Secretariat.

5.2 The 44th Annual Summer Institute in Survey Research Techniques

The Survey Research Center of the Institute for Social Research, The University of Michigan, will hold its 44th Annual Summer Institute in Survey Research Techniques at ISR during the summer of 1991.

Two four-week sessions, beginning July 1st and ending August 23 rd, will be offered. The program emphasis is on the sample survey as a basic measuring instrument for the social sciences. Teaching faculty in the Summer Institute are drawn primarily from the research faculty of the Departments of Sociology and Psychology from the University. Participants in the program gain familiarity with the application of survey research methods, including research design, sampling, measurement, questionnaire design, field methods, data management, and the statistical analysis of data. For a detailed brochure contact :

Dr. Duane F. Alwin, Director of the Summer Institute, Survey Research Center, The Institute for Social Research, The University of Michigan, P.O. Box 1248, Ann Arbor, MI 48016-1248. Telephone (313) 764-6595.

VI. COUNTRY REPORTS

6.1 Country report for Turkey by Dr. Oztas AYHAN

1990 Census of population

1990 Turkey Census of population was done on October 21, 1990 on de facto basis. Preliminary results are released on October 27, 1990. According to these results the population of Turkey is 56 969 109. The population of provinces and district centers are 33 666 967, and subdistricts and villages are 23 302 142. Final results are planned to be released on October 21, 1991. Inquiries about the Census results can be addressed to Census Division, State Institute of Statistics, Necatibey Caddesi n° 114, Ankara, Turkey.

Quality central survey of the 1990 Census of population

A quality control survey for the 1990 Census was done during the census field execution by independent teams. Survey is based on a stratified multistage sample design covering 19 provinces in 5 strata. A total of 443 enumeration districts are evaluated for coverage and response reliability. Survey results will be published with the Census results during 1991. Further information can be obtained from Research Division, State institute of Statistics, Necatibey Caddesi n° 114, Ankara, Turkey.

Baseline survey for health services to be implemented in 11 second priority provinces in Turkey

A baseline Survey is carried out covering 11 provinces in 2 regions of Turkey. Population is divided into 2 strata (urban and rural) and a sample of 3 000 households are selected by PPS. The fieldwork was carried out during mid October on de jure basis. Results will be released in a report form in the near future. Inquiries about the survey can be addressed to Director, Institute of Population Studies, Hacettepe University, Ankara, Turkey.

6.2 Country report for the United States by Dr. Daniel KASPRSZYK

The Institute for Social Research at the University of Michigan is designing and conducting a new Health and Retirement Study, in cooperation with the National Institutes on Aging. As currently planned, the first wave of data collection will begin in 1992, to be followed by biannual updates. The core sample for the study will be approximately 8 000 households in the coterminous U.S., with one or more members between the ages of 53 and 61 ; it is expected that a new cohort will be added in future years as the initial ages out of the lower end of the age range.

The new study has a number of unique features compared to other studies of retirement. First, there will be a massive effort during the year preceding the start of data collection to consult with academic specialists on retirement and aging : working groups including researchers from economics, medicine, health, gerontology, psychology, sociology, and demography will have major input into the survey content and other design features of the study. Second, there will be a substantial focus on health as it relates both to retirement decisions and to activity limitations in later life. And third, the study will pay a good deal of attention to changes in labor force participation that are likely to result in a higher degree of "jointness" to retirement decisions ; that is, substantial emphasis will be placed on understanding the family context of both retirement and family care decisions.

For more information about the study, correspondence can be addressed to : Dr. F. Thomas JUSTER, Room 3240, Institute for Social Research, P.O. Box 1248, The University of Michigan, Ann Arbor, MI. 48106.

A Quality Profile for the Survey of Income and Program Participation (SIPP), a large continuing national household longitudinal survey of the U.S. Census Bureau, has been prepared by Thomas Jabine with the assistance of Karen King and Rita Petroni of the Census Bureau Staff. The Quality Profile summarizes in convenient form what is known about the sources and magnitudes of errors in estimates based on data from the SIPP. While the monograph is obviously specific to the SIPP program, the authors' comprehensive approach to describing sources of error in surveys provides a useful framework for others who are conducting surveys. The SIPP Quality Profile is available at no charge from Customer Services, Data Users Services Division, Bureau of the Census, Washington, D.C. 20233 (301) 763-4100 or FAX (301) 763-4794.

Data Base News in Aging, is a newsletter that describes recent developments in U.S. governement sponsored data bases that focus on the older population. The U.S. Interagency Forum on Aging-Related Data produces the newsletter. The Forum consists of those U.S. Agencies that develop, collect, analyze, and disseminate data on the aging population. The Forum issues the newsletter 3 or 4 times per year. The newsletter is available for the cost of reproduction. For more information contact Arnold GOLDSTEIN, Population Division, U.S. Bureau of the Census, Washington, D.C. 20233 (301) 763-7883.

The monthly Product Announcement from the U.S. Bureau of the Census is designed to inform users about new products available form the U.S. Bureau of the Census. This monthly report contains a listing of publications and data files released by the Bureau during the previous month, announces upcoming products, and contains useful ordering information and telephone numbers. To subscribe to this free monthly report, contact : Customer Services, Data Users Service Division, U.S. Bureau of the Census, Washington, D.C. 20233 (301) 763-4100. Ask for a subscription to Monthly product Announcement.

The U.S. Census Bureau entered into a Joint Statistical Agreement with the National Opinion Research Center (NORC) to conduct a survey to measure factors that might be associated with the substantial drop in mail response rates in the 1990 U.S. Census of Population compared to 1980. The questionnaire included items directly connected with the census process, such as whether the household had seen the questionnaire, and mailed it back, and more general measures, including exposure to census publicity, attitudes toward government, surveys, and problems of junk mail and finding time. Analysis of the data, from 2478 responding households, is to be shared between the Census Bureau and the NORC. For more information, contact Robert FAY, DIR, Room 3067-3, U.S. Bureau of the Census, Washington, D.C. 20233, U.S.A..

6.3 Country report for the Philippines by Dr. Burton T. ONATE
Non-Sampling Errors

Studies on Non-Sampling Errors (NSE) are program components toward the improvement of the Philippine Statistical System (PSS). The National Statistics Office (NSO) and the Statistical Research and Training Center (SRTC) co-sponsored researches on NSE as applied to the Annual Survey of Establishments (ASE) and the Integrated Survey of Households (ISH) which are national sampling surveys on establishments and households, respectively. Results are also available on NSE in farm/holding survey.

These recent studies (1989/1990) developed the needed conceptual frameworks on NSE which include : coverage errors in the frame, unit and item non-responses, response or measurement errors, and processing and tabulation errors. These NSE have correspondence with the different universes and population parameters generated at different stages of the survey operations, namely : conceptualization of the target universe, frame development and construction, actual data collection (survey), and processing, tabulation and estimation procedures.

ASE, ISH and the Farm Survey identified separately the types and sources of NSE at different stages of the survey operations, developed measurements, set control limits for each type of NSE by strict supervision and effective management, and evaluated the overall accuracy of data for the information and benefit of users. This program is a new dimension toward the reduction of NSE ignorance in the country. (For more information, write to : Dr. Burton T. ONATE, P.O. Box 163, College, Laguna, Philippines, 40031).

6.4 Summary country report for Spain by Dr. Carmen ARRIBAS

Sample surveys :

- Annual industrial survey.

- Trade survey (each two or three years).
- Survey on the structure of restaurants, bars and cafeterias (1990 - first experience).

- Wages and salaries survey in industry and services (quarterly).

- Survey on hotels and campings occupation (monthly).

- Labour force survey (continuous survey, quarterly results at national, autonomous region and provinces level).

- Family budget survey :

a) Continuous FBS (sample size of 3 000 households - quarterly, follow up of structure of consumption at national level).

b) FBS - FACH 5 YFARS/28 000 households (up-date estimation of structure CPI, private consumption of national accounts...) results at national, autonomous region and provinces level.

- Survey on hospital morbidity (annual) 75 % of the hospitals - 10 % of deceases.

- Survey on the use of time of university professors (August 1989 - july 1990) - new experience, estimate ressources dedicated to R. and D.

6.5 Country report for Germany by Dr. G. FORBRIG.

Former German Democratic Republic :

Until 1989 in the German Democratic Republic only a few institutions dealt with sample surveys with respect to demographic, economic or sociological topics. Such institutions were e. g. the Statistical Office and the Institute for Sociology of the Academy of Sciences.

The most important survey of the Statistical Office collected data about income and expenses of households.

This survey included data of a year, but they were registered monthly. 3 400 households were included in this sample. Additional 400 households remained for three months in the sample. They were exchanged by other ones after this period.

Another survey reflected the different kinds of receipts and equipment of households. Equipment may be colour televisions sets, cars or houses.

A third survey reflected a balance of the daily time of people. The time was broken up into different groups. The balance showed how the time was used. The results were given for different groups of people e. g. with respect to the sex, to the age, the level of education, the number of children an other characters.

Since the beginning of 1990 experts from West- and Eastgermany prepared common activities for the future with respect to surveys. Since October 3rd, 1990 the Statistical Office of the Federal Republic is the organizer of official surveys. In the five countries of the former GDR a new statistical system will be introduced. There will be statistical offices of countries and of towns.

Instead of total collection of data more sample surveys will be organized, e. g. on price-statistics. Many other institutions and enterprises will deal with surveys e. g. on marketing and opinion research. It can be seen that the number of experts, who are interested in survey statistics, is increasing.

This year statisticians from the former GDR took part in the meeting of the German Statistical Society. Prof. STRECKER chaired the meeting of the Committee for Survey Statistics. The meetings of this committee are a good opportunity for the exchange of experience.

6.6 Country report for Italy by Mr. A. GIOMMI

Starting from the wave of July 1990, a new sampling strategy has been adopted by the Italian quarterly labour force survey.

The sample design is still a two stage stratified one, with communes as primary sampling units (PSU's) and households as secondary sampling units (SSU's). Changes regard the stratification method of the PSU's. The old method was based first on their demographic size in order to obtain two groups of communes : group A including those with more than 20,000 inhabitants, to be automatically selected in the sample (these communes are called self-representing : SR) ; group B of the other communes (not self-representing : NSR) which were successively strafified by altitude (mountain, hill, plane) and prevalent economic activity (agriculture, trade, industry). Only one commune from each stratum was to be selected in the sample.

The new stratification method is still based on demographic size in order to recognize SR communes. However, NSR communes are stratified simply by demographic size such that the resulting strata have approximately the same number of resident people. Finally, two communes are drawn from each stratum with probability proportional to their demographic size and without replacement.

Population parameters are estimated by means of combined post-stratified ratio estimator. Post-stratification is performed on the basis of both sex and age classes.

For more detailed information please refer to :

Dr. Aldo RUSSO
Istat (National Statistical Bureau)
Via C. Balbo n° 16
10100 ROMA - Italy

6.7 Country report for France by Anne Marie VESPA-LEYDER

The population of France in 1990 : results of the population census.

Starting on 5 March 1990 at 0 hours, INSEE carried out the 32nd census of the population in metropolitan France, the overseas departments (date of reference : 15 March), and the territories of Saint-Pierre-et-Miquelon. The first census was carried out in 1801 ; the previous census was conducted in 1982.

The census was carried out under the direct control of the Mayors. INSEE prepared and supervised the collection operations, and subsequently exploited the data. The Mayors recruited 110,000 census takers (1 for each 500 to 600 inhabitants), and INSEE provided over 3,000 delegates (specially recruited for the operation), who supervised the census agents after training them, and ensured liaison between the City Halls and regional INSEE offices. These delegates were supervised by 450 technical advisors from INSEE.

The results of the census revealed the number of households[1] and their composition, the socio-demographic characteristics of the inhabitants, as well as the number and comfort characteristics of the residences. Three questionnaires were used for the households : the individual form, the residence form, and the collective quarters file.

Apart from households, the census also took into account people in various communities, such as soldiers in barracks, elderly people living in retirement homes, members of religious communities, etc. These communities were directly censused by INSEE in the weeks preceding the date of reference for the household census.

The first estimates obtained on the basis of the counts of the questionnaires by census agents and City Halls started to be published at the end of June. The population of metropolitan France was 56,600,000 people as of 5 March 1990, as opposed to 54,335,000 on 4 March 1982, the date of the previous census. Thus, the annual growth rate

---

[1] Household : Group of people living in the same residence.

was 0,5 % which is close to that of the period between 1975 and 1982. Four fifths of this growth was due to an excess of births over deaths. The population of overseas departments was estimated to be 1,458,000 inhabitants in 1990, as opposed to 1,246,000 in 1982.

Demographic changes for 1975-1982-1990 have been published since September for each commune (of which there are 36,500 in France), cantons, departments, urban communities, and employment zones. The Order-in-Council legitimizing the results of the census (the legal population) will be published in December of this year.

More detailed structural data will be distributed starting at the end of 1990 and ending in 1993.

6.8 Country report for New Zealand by Dr. Stephen KUZMICICH

A pilot Time Use Survey (sample size 600), to assess the feasibility of measuring how New Zealanders spend their time, has been completed.

If the results of this survey are satisfactory and funds permit, a full survey will be implemented.

Major surveys approved recently include the subject areas of :

1. Research and development - information will be collected from businesses on expenditure related to research and development.

2. Balance of payments - expenditure of New Zealand travellers who are absent from New Zealand for less than 12 months.

Major Changes to Existing Survey

The sample size of the Household Labour Force Survey has been increased from 12,000 to 24,000. The increased accuracy provided will allow production of release of labour force statistics to be increased from quarterly to monthly.

The sample size of the Retail Trade Survey has been increased from 5000 to 8000 and other technical improvements have been implemented to improve the representativeness of the survey.

A census of companies with international financial investment transactions has been completed. The results from this census will be used to improve the methodology of existing Balance of Payments surveys.

To ease respondent burden on small businesses, an overlap control system has been implemented to prevent small Business from being selected from the department's central Business Directory for more than one of a series of business surveys.

Meetings and Conferences of Interest

A "statistics for Women" conference was held in August 1990. The conference discussed the relevance of statistics to topical issues of concern to women, and the adequacy of existing data on these matters.

A "Users and Suppliers Conference" is being conducted on 13 November 1990. Papers will be presented by staff of the department and by invited speakers. The conference will provide both users of statistics and suppliers of data with an opportunity for discussion on statistical issues.

Major Reports and Publications of Interest

1990 Yearbook. As part of New Zealand's 150th Anniversary celebrations the Department of Statistics changed the format of its Yearbook. As well as the usual statistics the 1990 Yearbook included 250 articles and 500 historical photographs illustrating facets of New Zealand's cultural, social and economic development since 1840.

The Fiscal Impact of Income Distribution 1987-88. This study examines the overall impact of government expenditure and revenue collection on the income of households in New Zealand.

The Human Face of New Zealand. This report describes the likely nature of population change over the next 60 years, identifying key population imperatives, then considers population policy in the context of the impact of these population imperatives on a range of social and economic policy sectors.

Elderly Population of New Zealand. This contains a variety of demographic information on the olderly population of New Zealand.

Economy Wide Census 1987. Seven reports will be produced covering the whole economy, each containing profiles of the industries in a major economic sector.

New Zealand Business Patterns 1989. A statistical analysis of New Zealand businesses, by industry, sector, employment, and location.

6.9 Country report for Canada by Dr. Gordon BRACKSTONE

Canada's next Census of Population takes place in June 1991. In developing the questionnaire for that Census, several cognitive research techniques, including observation of respondents completing the questionnaire, focus groups sessions, "in-depth" interviews, and "think aloud" interviews were used to investigate the extent to which respondents read and understood Census concepts, instructions and questions. As a result of this research many improvements to the Census

questionnaire were recommended, particularly in the design of the front cover and in the instructions that told respondents how to navigate through the questionnaire. A revised questionnaire that incorporated these recommendations was well received by respondents in a subsequent test and produced lower item non-response rates. This questionnaire became the model for the 1991 Census.

Another innovation for the 1991 Census is a full study to measure overcoverage. Since 1966, Statistics Canada has published estimates of gross undercoverage in the Canadian Census, based on a study known as the Reverse Record Check. In the 1986 Census, an experimental study of overcoverage was conducted in order to gain some experience with the methodology required. For the 1991 Census, the study is being refined and its scope is being expanded. The major component of the 1991 Overcoverage Study is a follow-up survey of some 30,000 private households that were enumerated in the Census to determine if each individual was in the target population for the Census, and if there were other addresses where the person might have been enumerated. The Census questionnaires for these alternative addresses will be checked to determine whether the individual was double-counted. Other components of the study focus on collective dwellings and on automated matching of individuals within local areas.

The objective of the 1991 Overcoverage Study is to produce reliable estimates of overcoverage for Canada and the provinces. When combined with estimates of undercoverage from the Reverse Record Check, this will permit the publication of official estimates of net coverage error in the Canadian Census for the first time. For more information on 1991 Census plans contact Don Royce, Social Survey Methods Division, Statistics Canada, Ottawa, Ontario, K1A 0T6, (613-951-6940).

Statistics Canada's annual methodology symposium in 1991 will be on the theme of "Spacial Issues in Statistics" and will be held in October or November 1991. Sessions will cover geographic considerations in the design, processing, presentation and analysis of survey data, and related issues of geographic infrastructure. For more information contact David Binder, Statistics Canada, Ottawa, Ontario, K1A 0T6 (613-951-0980).

VII. STATISTICAL ABSTRACTS

7.1 On estimating distribution functions and quantiles from survey data using auxiliary information

By J. N. K. Rao, Department of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6 ; J. G. Kovar, Business Survey Methods Division, Statistics Canada, Ottawa, Canada K1A 0T6 and H. J. Mantel, Departement of Mathematics and Statistics, Carleton University, Ottawa, Canada K1S 5B6.

Biometrica (1990), 77,2, pp. 365-75

Ratio and difference estimators of a population distribution function under a general sampling design are obtained, using auxiliary population information. The relative mean errors and relative mean square

errors of these estimators and a model-based estimator of Chambers and Dunstan (1986) are compared through a simulation study. The advantages of the design-based estimators over the model-based estimator under model mis-specifications, especially for large samples, are demonstrated. Ratio and difference estimators of a population quantile are also studied.

Some key words : Conditional properties, Difference estimator, Model-based estimator, Model mis-specification ; Ratio estimator.

7.2 Estimating the Size of a Subdomain : An Application in Auditing

By Lynne Stokes, Department of Management Science and Information Systems, University of Texas, Austin, TX 78712-1175.

J.B.E.S., July 90, vd 8, # 3, pp. 337-346

This article suggests an alternative to the ratio estimator for estimating the total size of a subdomain of a population. the application that served as the genesis for this work is from auditing. The problem is to estimate the total of sales transactions that are not tax exempt from an audit sample of the population of nontaxed sales transactions. A superpopulation approach, which models the unit's probability of belonging to the subdomain as a function of its size, leads to a family of estimators. The simplest member of this family is one in which that function is specified to be a constant. The optimal estimator for this model performs markedly better than the ratio estimator when the assumption is true and often performs better when it is not, though in that case it is biased. Stratification is shown to reduce this bias and at the same time make the ratio estimator more similar to the optimal estimator. A simulation experiment shows that the theoretical advantages hold in a real audit population.

Key words : Audit sampling ; Ratio estimator ; Superpopulation model.

7.3 Small-Area Estimation of Economic Statistics

By Cary T. Isaki, Statistical Research Division, U.S. Bureau of the Census, Washington, DC 20233.

Journal of Business and Economic Statistics, October 1990 Vol 8, n°4, pp. 435-441.

A ratio-correlation (multiple regression) approach for estimating key economic statistics for small areas is proposed and compared with several synthetic estimation methods using various measures of performance. Using published data that are easily accessible. The methods provide a means of estimating statistics not currently published with a reasonable amount of error.

Key words : Country Business Patterns ; Mean absolute relative error ; Multiple regression ; Ratio-correlation methods : Synthetic estimation.

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.

24.1 Question : The terms truncate, censor, trim, and Winsorize, among others, seem to be used interchangeably and confusedly in referring to the handling of "outliers" in study data. What do you recommend as to their usage in survey data?

Answer. (Contributed by Prof. Lawrence H. Moulton, Biostatistics Dept., SPH, University of Michigan, 48106).

You are correct in that many of these terms are rather loosely bandied about. The first thing to recognize is that they fall into two distinct classes: 1) truncation and censoring are characteristics of the study design, determining whether values of certain variables would be known to or used by the investigator; 2) trimming and Winsorization are statistical analysis techniques that one applies (as a transformation) to a set of data once it has been collected.

Data are referred to as truncated when they are sampled from a conditional distribution, so that the total number of data points is unknown. For example, if one were trying to estimate the mean length of fish in a pond, but only fish of a certain minimum length could be caught in a net to be measured, the population would be said to be truncated. In the sample survey context, characteristics of units may cause them to be outside the sampling frame, and hence their response values truncated, as in a survey of "business establishments with at least ten full-time employees." This may be viewed as lack of coverage of the sampling frame.

Censoring, on the other hand, occurs when a respondent has been identified, but the precise value of their response value is unknown because it has exceeded or failed to exceed a given value. This is common, for example, in estimation of lifetime distributions, where an individual's exact survival time may be censored by loss to followup or by living past the end of the study period.

The term "outlier" often refers to a value which is so extreme compared to the rest of the data that it may well be considered as having emanated from a distribution different from the one under investigation. Thus, data may be "contaminated" by outliers. Ways this can happen include inclusion error, as when a millionaire is included in a sample of supposed public aid recipients, or from erroneous data, resulting from digit transposition, deception, etc. However, "outliers" may also represent extreme values from skewed distributions (e.g., incomes of millions or billions), which are excluded (beyond some stated limit) because they would unduly increase variances and especially variances of variances.

Trimming or Winsorizing the data can yield an analysis that is robust to the assumption of "no contamination by outliers" that is often made. A trimmed analysis results from analyzing the data which remains after throwing out a certain proportion of the smallest and largest data values, while a Winsorized analysis proceeds by first replacing the most extreme values by the closest order statistics. For example, the 20% trimmed mean of the data: 35, 51, 52, 56, 98 is calculated as: (51+52+56)/3 = 53, while the corresponding Winsorized mean replaces 35 by 51 and 98 by 56 to get: (51+51+52+56+56)/5 = 53.2 . Asymmetric application of these methods also may be appropriate. Trimming generally is preferred to Winsorizing because of the latter's overdependence on the particular values of the next most extreme values, 51 and 56 in the preceding example.

In the sample survey setting, if one is interested in a mean of a variable, and it is acknowledged that error or contamination of the data may be present, trimming the data can be highly efficient. However, in many survey settings it is possible to return to verify the data, in which case it may be preferable to allocate resources to routinely check, say, the largest and/or smallest 5% of the observations.

Editor's note: Curtailing is also used; and shrinking and transformations may also be considered alternative methods for treating these problems. For references see the new Wiley Encyclopedia of Statistics and the new Longman's Dictionary of Statistical Terms by F.H.C. Marriott.

24.2 Question : Problems of estimating the differences between two means or two proportions from sample surveys often arise with correlations, covariances between the pairs of means. The covariances can be either between two subclasses of the same survey, or between similar means from two periods, or from a panel survey. How do you deal with the covariances in estimating variances of the differences?

Answer. These problems have been treated in various forms for data from sample surveys. Fortunately the variances of differences of ratio means $(r_c - r_b)$ can be computed with the same methods as the variances for ratio means $r = y/x$, as shown with computing formulas and numerical examples in [Kish, Survey Sampling, Wiley 1965, Section 6.5].

The same formulas will deal with the complexities from panels of the same individuals, or from merely overlapping area segments, or from only reuse of the same secondary units (blocks), or only from reuse of the same primary units. Of course, the effects of positive correlations and covariances, hence reductions of the variances, to be expected for most variables in $\mathrm{Var}(r_c) + \mathrm{Var}(r_b) - 2\mathrm{Cov}(r_c, r_b)$, will be least for the primary units and greatest for panels. These can be written generally as $2S^2(1\text{-}R)$, and the correlation R is greatest for panels. Often $R > 0.5$ and then the variance of the differences can be less than the variance S^2 of single means. When the overlap is only partial in the portion W of the sample the variance is $2S^2(1\text{-}WR)$. These matters are discussed in [Kish, Section 6.6C, 14.1, 12.4-5], and elsewhere. Proportions p are special cases of means and need no separate treatment from r when dealing with complex survey samples.

For proportionate stratified element sampling the covariance terms tend to vanish, and the variance of the difference tends to the sum of the variances. Furthermore, the gains due to proportionate stratification tend to vanish for small subclasses. Thus the variance of the difference tend to the srs formulas [Kish, 1965, Section 4.5A].

The formulas for differences are useful because the same formulas apply to:

1. Differences of means (r_c-r_b) between two subclasses of the same survey from the same primary selections.

2. Differences of two means (r_c-r_b) from two periodic surveys from the same primary selections; and secondary selections, and overlapping (partial?) segments.

3. Differences of two measurements made on the same panel, either between two periods, or as a "test-retest" or "before-after" situations.

4. Matched pairs of sampling units with similar problems of covariances.

5. Differences between two categories of a polytomy $(p_1, p_2, \ldots p_k)$ also have covariances that can be treated similarly.

6. The sums of three ordinal scale values 0, 1, 2 (or k, $k+1$, $k+2$) also have similar treatments, surprisingly.

These matters take special forms for simple random samples (srs), when $\mathrm{Var}(p) = PQ/n$. Then $\mathrm{Var}(p_c-p_b) = [P_cQ_c + P_bQ_b - (P_c-P_b)^2]/n$, the third term representing the covariance, curiously. These are shown in [Kish, Section 12.11], with references to earlier demonstrations and uses. This problem for srs gets repeated attention in the literature, but without mentioning the srs assumptions, or the earlier treatments, and often dealing with "exact" significance tests. The latest I found is by A.J. Scott and G.A.F. Seber in the American Statistician, 1983, 37, 319-20. However, it is misleading to advocate these for survey proportions, because design effects from clustering may be present and not negligible.

IX. PUBLICATIONS

A Dictionary of Statistical Terms : Fifth Edition

The publisher of this Dictionary, Longman Academic Scientific and Technical has agreed to offer the ISI and its sections a 20 % discount on its purchase price.

Additional information is posted in the November edition of the ISI newsletter.

Members may place orders by contacting the ISI Permanent Office in Voorburg, The Netherlands. The purchase price (normally £ 25) is now £ 20, 60 D.M., 36 US $, 200 F.F., 67 D. fl. or 51 S.F.

Orders should include payment and should indicate the number of copies desired.

INSEE.D320.0056.90.0.

SURVEY METHODOLOGY

A Journal of Statistical Development and Applications

Survey Methodology publishes articles dealing with various aspects of statistical development such as,

* design issues in the context of practical constraints
* use of different data sources and collection techniques
* total survey error
* survey evaluation
* research in survey methodology
* time series analysis
* seasonal adjustment
* demographic studies
* data integration
* estimation and data analysis methods
* general survey systems development

Each article focuses on developing and evaluating specific methodologies for data collection or data evaluation.

In addition to general topics of current interest to survey statisticians, most issues of **Survey Methodology** contain a special section with a concentrated treatment of new techniques and experiences for a selected topic.

All articles in **Survey Methodology** are refereed by an international board and the journal enjoys world wide circulation as a result of cooperative arrangements with various statistical associations.

Invitation to Authors - Authors are invited to submit manuscripts in either English or French. For more information, please write to: Editor, Survey Methodology, Methodology Branch, Statistics Canada, Ottawa, Ontario, Canada, K1A 0T6.

The regular annual subscription rate to **Survey Methodology** is $49US. However, members of the IASS receive a substantial discount when they subscribe through the IASS.

Editorial Board: Editor - M.P. Singh. Associate editors - B. Afonja, D.R. Bellhouse, D. Binder, E.B. Dagum, J.C. Deville, G.D. Drew, W.A. Fuller, J.F. Gentleman, M. Gonzalez, R.M. Groves, D. Holt, G. Kalton, J.N.K. Rao, D.B. Rubin, I. Sande, C.E. Sarndal, W. Schaible, F.J. Scheuren, K.M. Wolter.
Assistant editors: J. Gambino, L. Mach, A. Théberge.
Management Board: G.J. Brackstone, N. Chinnappa, G.J.C. Hole, C. Patrick, F. Mayda, R. Platek, D. Roy, M.P. Singh.