INTERNATIONAL STATISTICAL INSTITUTE
INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

NEWS LETTER Nº 8

SURVEY STATISTICIAN Nº 0

OCTOBER 1978

CONTENTS

EDITOR'S LETTER

IASS Newsletter nº8 is published in a new way. The pattern of the Newsletter of the International Statistical Institute "International Statistical Information" has been adopted to make the reading of the Newsletter more pleasant and also... to cut mailing costs.

Two versions are published : an English one and a French one.

The Newsletter is too "the Survey Statistician nº0", the publication that will replace the Newsletter from now on. This is only a draft and the Editing Board would be happy to receive suggestions and critics from the members of the Association about the pattern of this publication and the contents of its sections. We will as much as possible take in account the observations received to prepare the Survey Statistician nº1 that should be issued in March 1979.

GENERAL INFORMATION

WORKSHOP IN SURVEY SAMPLING, MEXICO, NOVEMBER 6-10, 1978

A Workshop in Survey Sampling will be held in Mexico City on November 6-10, 1978. This will be a regional workshop for Latin America, conducted in Spanish, and in accord with recommendations of the Council of the IASS for Regional Workshops. Participants are asked to bring actual, practical, unsolved problems for joint discussions and solutions.

In addition several internationally recognized survey statisticians will participate and promote discussions

in four areas :

- sampling in censuses
- small area statistics
- construction and maintenance of master sample frames
- nonsampling errors

For more information write to :

Hector HERNANDEZ-Llamas
Universidad Autonoma Metropolitana-Xochimilco
Calzada del Hueso y Canal Nacional
Apdo. Postal 23-181
Mexico 23, D.F.

or Edmundo F. BERUMEN
Eugenia 1202
Col. del Valle
Mexico D.F.

We should like to see more IASS Regional Workshops elsewhere also.

Leslie Kish
Workshop Committee

PROPOSITIONS FOR MEETING TITLES FOR IASS 1979 MANILA SESSION

Preliminary titles	Proposed organizers
Technical information exchange on statistical software (TIESS) with emphasis on survey and census processing	I. Francis (New Zealand)
Interactive computing, interactive procedures with large data files	I. Francis (New Zealand)
Rotation and other resampling schemes	I.P. David (Philippines)
Incomplete data	W.G. Madow (USA)
Estimation of population growth	W. Brass (UK) & G. Calot (France)
Enterprise statistics	G.K.G. Forbrig (G.D.R.)
Privacy and confidentiality issues in surveys	T.B. Jabine (USA)
Market intelligence surveys	S. Tulya-Muhika (Uganda)
Surveys in developing countries	M.N. Murthy (India)

SURVEY CENTERS IN AUSTRALIAN UNIVERSITIES

There are three academically based Centres specializing in sample survey work in Australia. These are the Survey Research Centre at the Australian National University, the Sample Survey Centre at the University of Sydney, and the Centre for Applied Social and Survey Research at the Flinders University of South Australia.

The activities of these three Centres are to a large extent complementary and between them they offer a wide range of services to social scientists both in their own Universities and outside.

1. Australian National University, Canberra

The Survey Research Centre at the Australian National University is primarily intended to service the University's various social science research sections, but it is also co-operating with other Universities in establishing and maintaining facilities for social scientists in general.

The Centre provides a consultation service for persons and organizations conducting or wishing to conduct surveys, and provides advice and assistance on request. The Centre also provides survey processing facilities for ANU Departments.

From March 1978 the Centre is conducting a six-monthly Canberra Population Survey, which is primarily intended as a service to academic and government departments. Interviewers call at unclustered sample of over 500 dwellings throughout the Canberra city area. Questions of topical interest are asked, in addition to those sponsored by outside organizations. Clients have the option of receiving clean data tapes or tabulations based on their own specifications.

The staff of the Centre are engaged in research into the statistics of survey work, in particular sampling theory and the analysis of repeated surveys. They conduct a graduate course in sample survey theory and practice.

Facilities are also available for students to do postgraduate work in the Centre.

The Centre provides the secretariat for the Australian Consortium for Social and Political Research Inc. (ACSPRI). This is an association of twelve Universities and Colleges of Advanced Education which holds the Australian National Membership of the Inter-university Consortium for Political and Social Research (ICPSR). Through this Consortium, social scientists have access to a large number of thoroughly edited data sets, facilities for subsidized attendance at summer programs, and assistance with computer processing of surveys. As a by-product of this activity, the Centre is gathering a collection of data sets for ACSPRI. It is also assembling a catalogue of these data sets, an inventory of Australian surveys, and a listing of relevant journal abstracts.

The current staff of the Centre consists of a Director, Research Fellow, Research Officer, and three supporting staff.

2. University of Sydney

The Sample Survey Centre at the University of Sydney is primarily concerned with the needs of that University's staff, but a limited amount of assistance is available to persons or groups outside the University. This does not apply where any data collected is to be used for commercial or private purposes.

The Centre maintains an apparatus for the conduct of all types of surveys of the population, performs basic research into the theory and practice of surveys, and assists in the teaching of courses involving survey methods.

All phases of survey production can be handled by the Centre including sampling, questioning strategy, questionnaire design, field interviewing and supervision, coding, editing, computer analysis and reporting. Users may avail themselves of all or some of these services.

The Centre also maintains its own data base -including extensive tape, microfiche, hard copy (and map) holdings of data collected by the Australian Bureau of Statistics- and can arrange ready access to survey data base collections houses elsewhere in the world.

The Centre has a small permanent staff of academic and technical personnel including a Director, Survey Research Officer, a Programmer and Research Assistants. It engages part-time field and office staff to meet the specific demands or projects commissioned by clients.

3. Flinders University of South Australia, Adelaide

The Centre for Applied Social and Survey Research (CASSR) was established by the University Council in 1977 to promote applied social and survey research, and to disseminate the findings of research through the publication of technical research papers and monographs and the holding of seminars, workshops and summer schools.

In addition to conducting its own research activities and providing a survey research facility for staff and graduate students within the University, CASSR can provide consulting advice to government and private enterprise in a wide range of applied research fields and conduct sample survey research and data analysis in accordance with the University's outside research contract policy.

The principal aim of CASSR is to develop its own academic research activities involving staff drawn from disciplines within the School of Social Sciences. Such projects are dependent on external funding. Areas of current research include the planning and utilization of health care services, location planning, market research, Aboriginal studies, problems of living in isolated communities, population projections, social indicators, unemployment in country towns, voting behaviour, spatial data banks, and the design of safety symbols.

A major research emphasis of CASSR is in the area of survey design, management and data analysis and interpretation. It maintains a probability sample of households in metropolitan Adelaide and other areas of the state, and in collaboration with similar centres interstate can conduct survey research outside South Australia. A panel of interviewers is on call throughout the year. The Centre has access to the University's DEC-10 computer, has its own on-line terminal facilities, and is developing a wide range of software for the analysis of survey research data.

CASSR's activities are controlled by a management committee which comprises a Director, academic staff of the University, and a number of experts in social research from outside organizations. People involved in the research activities of CASSR may be appointed Research Associates of the Centre. It has a full-time research assistant, as well as clerical and secretarial staff of 6.

Further information on each of these Centres can be obtained by writing to the appropriate Secretary. The addresses are :

ANU Survey Research Centre
P.O. Box 4
Canberra, A.C.T. Australia 2600

Sample Survey Centre
University of Sydney
N.S.W. Australia 2006

Centre for Applied Social and Survey Research
Flinders University of South Australia
Bedford Park
South Australia, Australia 5042

from Dennis TREWIN,
Australian Correspondent

SECOND CONFERENCE ON SAMPLE SURVEYS, MELBOURNE, 28-30 SEPTEMBER, 1977

The second conference on Sample Survey Theory and Practice was held in Melbourne. There were 85 participants of whom about 30 were from CSIRO and the rest from Government Departments, Universities and other educational institutions and industry. Dr Gani, Chief of the Division of Mathematics and Statistics opened the conference and

stressed the importance of sample surveys and mentioned briefly some examples in his opening remarks. This was followed by Ken Brewer's (Survey Research Centre, ANU) talk in which he presented a sample design appropriate for large scale enterprise and establishment surveys. Next Gary Morgan (Morgan Research Centre) discussed the weakly Australia-wide "omnibus" survey with special reference to the problem of "not-at-homes". Phil Hughes (ABS) outlined the sample design for the quarterly and monthly retail surveys and David Cocking (ABS) discussed the use of sampling procedures in the 1976 census of population and housing. The post lunch session started with a paper by T.J. Rao (DMS) who presented some alternative estimators of change in proportions over successive occasions. Their use in market research surveys and opinion polls was mentioned. This was followed by a paper on marine pollution study presented jointly by Noel Coleman (Marine Pollution Study Group) and Wilf Cuff (CSIRO, DCR). They presented the details of a benthic survey conducted in Western Port Bay during 1973-74 with a view to assessing the environmental conditions. The final talk was on "At-the-scene study of road accidents to which an ambulance is called in metropolitan Adelaide". This was again jointly given by Jack McLean (Road Accident Research Unit, University of Adelaide) and Geoff Robinson (Australian Road Research Board).

Leon Kempen (ABS) started the second day's proceedings with a paper on the application of grouped data methods in the analysis of household expenditure survey. This was followed by the joint paper by David Ratkowsky (DMS) and Elspeth Longmore (DMS) on Tasmanian child growth study. It was presented by David Ratkowsky who discussed several surveys on child growth conducted in Australia, New Zealand, U.S.A. and the U.K. and compared them. Gary Dickinson (Applied Mathematics Division, DSIR, New Zealand) discussed the application of general linear modelling to the analysis of contingency tables arising out of data collected in the Magistrate's Courts. The aim of the analysis was to test the effect of such factors as representation, race, offence, appearance from custody or bail, objections to bail and whether remand was for trial or sentence, on the decision to remand in custody or bail. Next Allan Nicholls (ABS) outlined how regression models can be used to estimate the relationship between population growth and certain "symptomatic indicators" during the intercensal period. After lunch, Gordon Mc Clatchie discussed the problems of cross tabulation in epidemiological surveys when multi-valued data items are present. This was followed by the paper on a two-phase psychiatric community survey conducted by the Social Psychiatry Research Unit at the ANU, presented by Paul Duncan-Jones (ANU), which examined relationships between social factors and psychiatric morbidity. Next Alan Jones and Graham Mascas (ICI Australia) posed the problems of using data collected from a voluntary survey conducted the Sydney Coronary Heart Disease Prevention Programme.

The final morning started off with a talk by Phil Hughes (ABS) on the use of collocated sampling in the Bureau. Allan Nicholls (ABS) then outlined the details of the survey of employee earnings and hours. This was followed by David Cocking's (ABS) description the Generalised Survey Design and Processing Systems in the ABS. Next John Kerr (DMS) presented a paper relating to a survey of acquired resistance to chemicals in the cattle tick "Boophuüs Microplus". This survey was undertaken by the CSIRO Division of Entomology with the cooperation of the ABS and the Queensland Department of Primary Industries and was designed to sample cattle owners of the tick-infested areas of Queensland. The post lunch session consisted of a paper on estimating life of power poles in Victoria by Frank Arundell (State Electricity Commission of Victoria), followed by John Beesley's (CSIRO, DBR) paper which outlined six sample surveys initiated by the wood preservation section of the old CSIRO Division of Forest Products. This talk was illustrated with interesting slides.

from T.J. RAO
DMS Newsletter, CSIRO Division of Mathematics and Statistics
N° 39 January-February 1978

MEETING ON THE PROBLEMS CONCERNING HOUSEHOLD SURVEYS, GENEVA, MARCH 20-23, 1978

Among the papers presented, the following ones may be quoted :

- Problems raised by the optimal allocation of funds at each step of a survey (Sweden)

We discussed mostly on the advantages of automatic coding :

- reduction in time : this is determining when the publishing time must imperatively be short
- reduction in cost : even if the preparation (acquisition) cost of a code dictionary is included
- lack of variability due to human intervention

Yet the method does not allow to cancel the document checking before coding : it is meant to find out missing data, which are corrected in a manual or automatic way.

Therefore, automatic coding should not be considered as a method to be practised separately but to be integrated into a processing system.

An inconvenient has been pointed out in one country : the variety of languages within a same country prevents the use of one stored dictionary and therefore makes more difficult the automatic coding implementation.

- Problems raised by missing data in a survey processing (Canada)

The meeting dealt with the philosophy and the various concepts of imputation(1).

During the discussion, the Canadian delegate was brought to precise that the comparison between imputation and a posteriori correction methods, gave variable results concerning the quality of the results ; he had no data on cost and time considerations.

- Methods of reduction in non responses (U.K.)

The paper dealt with the problem raised by non responses. A clear and detailed analysis was given on this complex issue.

(1) In the paper, imputation is defined as the allocation of data to vacant zones (including full questionnaires) or the replacement of non valid data, according to a series of rules

During the discussion, the Canadian representative spoke of the studies made on the matter in Canada :

- non response rate with respect to some interviewer characteristics : no determining factor could be stressed out
- comparison between known respondents and non respondents characteristics : some results are confirmed (difficulty to get in touch with small-sized households), others remain unexplained (difficulty to contact unemployed people)

Canada developed interviewer training programs together with methods to assess those programs, which allows to appreciate the work done, objectively.

The Swedish delegate then talked about the results of some works realized in Sweden, principally the distinction of four groups of non respondents :

- sick, old people
- people suspecting a dishonest use of the provided information
- people very busy with their family occupations
- people showing no interest for surveys

The French delegate mentioned the possibility, if the survey allows it, to interview a household later than the reference date, which however introduces a survey bias. He also talked about the "contagious" effect of refusal, especially in small "communes" where some negative influences can appear.

At last the representative from the United Kingdom pointed out the attitude of the OPCS[4] towards the advertising of surveys : there is a clear opposition to notice letters owing to experiments quite negative on the matter ; nothing can replace a well-trained interviewer to get in touch with people.

- Among other reports, we should point out the paper presented by the representative of the World Fertility Survey ; it dealt with a software meant to determine random errors in cluster sampling

Let us mention too the papers from Eastern Europe countries that, apart from the Yugoslavian paper described "Family budgets" surveys, focusing above all on sampling designs and on the desired interviewer behaviour.

During the final discussion, two directions appeared :

- the study of recommendations on the presentation of the errors affecting the published results, so that users know the result reliability. The matter would be to focus on the problems of definition of concepts, updating, coverage errors, sampling, non responses, processing together with the problems of comparibility with other sources.
- a larger exchange of information on the experiments implemented by the participating countries.

---

(1) Office of Population Censuses and Surveys

from J.M. REMPP and G. ROY
INSEE, France

SURVEYS IN PROGRESS

SURVEYS LED BY THE AUSTRALIAN BUREAU OF STATISTICS

A - 1977 Surveys

1. Quarterly Labour Force Survey
2. Supplementary surveys included on the Quarterly Labour Force Survey were :
   - i Labour Force Experience
   - ii Internal Migration
   - iii Tobacco Consumption
   - iv Alcohol Consumption
   - v Persons looking for work
   - vi Persons not in the Labour Force
   - vii Child Care
   - viii Multiple jobholders
   - ix Frequency of Pays and Weekly Earnings of Wage and Salary Earners
   - x School Leavers
3. Quarterly Survey of Job Vacancies
4. Survey of Weekly Earnings and Hours
5. Survey of Employee's Earnings and Hours
6. Survey of Tradesmen and Apprentices
7. Survey of Local Government Employees
8. Survey of Stocks
9. Survey of Capital Expenditure
10. Monthly Retail Survey
11. Quarterly Retail Survey
12. Survey of Interest, Rent and Royalties
13. Survey of Research and Development Expenditure
14. Quarterly Survey of International Trade Credit
15. Quarterly Survey of Import Values
16. Monthly Survey of Import Orders for Sensitive Commodities
17. Income Tax Survey
18. Combined Cereal Surveys
19. Survey of Sheep and Wool
20. Survey of Agriculture Finance

B - Some recent surveys conducted by Australian Bureau of Statistics (ABS)

1. Starting in February 1978, the ABS's Quarterly Population Survey has become monthly with the emphasis on labour force statistics.

2. The South Australian Office of the ABS has assisted the South Australian Royal Commission into Non-Medical Use of Drugs has successfully conducted a Survey of Non-Medical Use of Drugs on 1% sample of dwellings in Adelaide (population approx. 900,000)

3. The New South Wales Office of the ABS will shortly be conducting a Housing Survey in the Statistical Areas of Sydney, Newcastle and Woolongong. The survey will be of about 8, 000 dwellings. It will provide information on the underlying reasons for the changing demands for housing, employment and facilities and also the reasons for the patterns of intra- and inter-regional migration.

4. The ABS conducted a Survey of Interest, Rent and Royalties paid by business enterprises. The survey used the same framework as the Quarterly Survey of Capital Expenditure, and the framework is derived from the ABS's integrated register of business establishments, enterprises and enterprise groups.

5. The ABS conducted a survey of Research and Development Expenditure. The survey used multiple frame methodology in that it used names and addresses obtained from the Australian Industrial Research and Development Incentives Board to provide part of the sample and a framework of business enterprises derived from the ABS's Integrated Register for the remainder of the sample.

6. The ABS's first Health Interview Survey is nearing completion. The survey design incorporated the joint objectives of providing accurate regional and national objectives. It used a stratified multi-stage area sample.

from Dennis TREWIN
Australian correspondent

PAPER SUMMARIES

OPTIMAL ASSIGNMENT OF MONITEURS (INTERVIEWERS) APPLIED TO OFFICIAL AGRICULTURAL STATISTICS IN BELGIUM

During the survey of agricultural statistics in Belgium in May 1968 conducted as a total survey by the Institut National de Statistique (I.N.St.), Belgium, 51 moniteurs (interviewers) of the I.N.St. went to 269 villages selected by random to check the answers of the interviewed persons. The total time of the field work of such a survey could be reduced considerably if the moniteurs would be assigned optimally to the villages with regard to their itineraries.

An optimization model has been set up applying methods of operation research and optimization theory to that purpose. This has been the first attempt to introduce such methods into official statistics in Belgium. The feasibility to determine the moniteurs' itinerary in order to save time and money and to reduce the total time of the field work has been checked. The moniteurs are living in different places, e.g. Brussels, Antwerpen, Liège, Mons, etc., and have to travel to villages spread all over the country to carry out surveys.

Taking into consideration different working conditions for the moniteurs (eight-hour-day or spending the night on the road or not) the total time of the field work has to be minimized. Other targets e.g. minimizing the costs, can be optimized alternatively. The whole question can be considered as a "non-linear transportation problem", including a "multidimensional-traveling-salesman-problem". A non-linear optimization model has been developed including a sequential algorithm to achieve approximative solutions. This model was applied to the data of the survey of agriculture and horticulture in Belgium, May 1968, I.N.St.

Comparing the empirical data with the calculations based on the optimization model, the travelling time has been reduced by about 35 % in the Flamonian and about 50 % in the Wallonian region. The total time of the field work was reduced by 41 % and 33 % respectively.

Other criteria, e.g. the total number of interviewers required, showed similar improvements.

Another advantage of this model is that the division into subregions once made for this survey, can be used for later surveys too.

Furthermore, the application of this optimization model reduces also certain kinds of non sampling errors.

The whole problem was analyzed in cooperation with I.N.St. ; it may be considered as a first approach to apply optimizing methods to official statistics. This model is certainly not perfect. The authors will be grateful for all critics and proposals for improvement.

R. WIEGERT, K. KAFKA, H. STRECKER, R. STEYLAERTS
INSt, Belgique, Bulletin de Statistique n° 1-2, janvier-février 1977, 63e année, pp. 37-59

LABOUR FORCE SURVEY SAMPLE SIZE STABILIZATION

In the past, the LFS sample size (in terms of households) has grown at the rate of 2-3% per year. This increase was due to the natural population growth, a declining average household size, and due to constant sampling rates applied within the LFS. The sample size stabilization method was developed in order to control the size of the LFS sample at a pre-determined level. By restricting the growth in the number of sampled households to that necessary to maintain a constant number of sampled persons, the reliability of survey results will be unaffected, while the expense of data collection and processing can be reduced.

In order to describe the method, background on some aspects of the LFS design dealing with the selection procedures may be helpful. The LFS utilizes a multi-stage self-weighting design in which the next to last unit of selection (the cluster) corresponds to a land area having from 3-100 dwellings, depending upon the type of area. At the time of design, each cluster is assigned an inverse sampling ratio (i.s.r.) which specifies the rate at which the cluster would be sub-sampled, should it be selected, to ensure adherence to the overall basic sampling rate. When clusters are initially selected, the cluster i.s.r. (as step interval) and a random start(between 1 and the cluster i.s.r.) determine a systematic sample of dwellings. Every six months, rotation of households within a cluster is effected by increasing the random start by 1, which results in a new set of dwellings. When the incremented value of the random start exceeds the cluster i.s.r., then rotation takes place into the next cluster in the frame with a random start of 1. At the cluster level, the LFS sample is divided into 6 representative parts, called rotation groups, which govern the month in which sample rotation occurs, i.e., rotation 1 in January and July, 2 in February and August, etc. Updated dwelling lists are stored on computer for selected clusters, allowing automated selection of dwellings each month for the 1/6 of the sample which is rotating. There may be new dwellings, however, which are added to the list by the interviewer after computerized selection for a given month. These dwellings are sub-sampled by interviewers at the same (within-cluster) rate.

The method for stabilizing the sample size involves systematically dropping, across areas within province sharing the same basic sampling rate, a subset of the dwellings newly selected by computer each month in order to maintain the sample size at a pre-determined level. At the estimation stage, a compensating weighting factor (the sample stabilization weight) is applied separately for each rotation group within these sub-provincial areas.

It should be noted that dwellings falling into three categories are not subjected to the drop, and hence receive a sample stabilization weight of one :

(i) The new dwellings manually selected by the interviewers. At the processing step where manual selections are verified as being valid, they are flagged so that throughout their life in the sample, a sample size stabilization weight of one will be applied to records from these dwellings.

(ii) An open ended apartment frame exists in larger cities. If a cluster (apartment buildings) added to the frame is selected, then it is introduced into the sample in three months time regardless of its rotation number. If introduced "off rotation", then the originally selected dwellings will not remain in the sample for the full six months, as rotation will occur in the usual month governed by the rotation number. The off-rotation clusters are not subjected to the drop. A file of off-rotation clusters is generated each month from the dwelling selection program for input to the wieghting system.

(iii) In order to avoid disruptions to interviewer assignments, any cluster experiencing exceptional growth since the time of design (400%-500%) is generally further sub-sampled by changing the i.s.r. to a multiple of its original value and applying a compensating cluster sub-weight at the estimation stage. The basic rate at which these clusters are sampled will be different from the rate at which regular clusters are sampled, hence these dwellings are excluded from the drop. A file of sub-sampled clusters is input to both the dwelling selection program and the weighting system.

The sample size stabilization feature is being introduced into the LFS rotation by rotation over the period November 1977 to April 1978 in order to stabilize the sample size at its March 1977 level of 56,600 households per month. The method is flexible in that a change in the sample size can be effected by changing a table of prespecified or base sample sizes which are stored in the dwelling selection program.

The sample size stabilization feature can also be extended to provide a flexible sample selection capacity for supplementary and special surveys operating on the LFS frame. Particularly for supplementaries it will permit sub-sampling the Labour Force Sample by area types (NSRU urban, NSRU rural, SRU, apartments, special areas) within provinces, at any rate from maintaining a full LFS sample to exclusion of the area. For special Surveys, it will also enable greater optimality criteria to be utilized in sample allocation whereas in the past, within area type sub-provincial allocations usually had to conform to a number of complete rotation groups, with the sample stabilization, the rotation groups can be sub-sampled to arrive at an optimal size. Of course, for the purposes of budgetary control, it will be a great advantage for both types of surveys to be able to pre-specify to exact sample size. In addition, the dwellings that are dropped each month are themselves representative samples of dwellings. These may if found desirable, be used for conducting small tests or surveys at the national level.

Enquiries relating to the LFS Sample Size Stabilization should be directed to Dough Drew, Household Surveys Development Division, Statistics Canada.

from "Quarterly Bulletin of the Methodology Divisions
Statistical Services Field, Statistics Canada"
N° 25, January 1978

QUESTION/ANSWER

PRESENTATION OF THE SECTION

In a recent IASS Newsletter, Professor Kish made the interesting suggestion to open a Question/Answer section. On our request, Professor Kish willingly gave the following precisions.

This Q/A Section of the Survey Statistician is devoted to practical problems of survey sampling of general interest. We invite you, the readers,to send in practical problems you encounter in your actual work in survey sampling. We intend to publish answers to those which seem to be of general interest. Questions that do not have those criteria, and questions for which published answers are readily available will be answered with private letters from the editor or from a selected expert on the topic. You will receive friendly answers in any case.

For confidentiality, we shall omit the name and hide the source and country of the questioner, except when permission is clearly indicated. But for the sake of further reference, the answers will carry names, and not always the editor's (Mr Kish's) because he expects to find experts for special topics.

Send your questions to the IASS Secretariat, c/o INSEE, 18 boulevard Adolphe Pinard, 75675 Paris, Cedex 14, France and to this section's editor : Leslie Kish, Institute for Social Research, The University of Michigan, Ann Arbor, Michigan 48106. But if you indicate that you sent it only to one of those, a copy will be sent to the other.

We begin with a question synthesized from several concerned with exhausted units and submitted to the IASS Workshop on Practical Problems of Survey Sampling at Delhi, December 1977. It seemed interesting and useful to the thirty statisticians there, hence it should also be to other readers.

J. DESABIE
Editor of the "Survey Statistician"

EXHAUSTED AND UNDERSIZED PSU's

Question :

Our national sample of administrative districts was selected with PPS (probabilities proportional to size in the 1970 population Census) several years ago. They have served as primary selected units (PSU's) for several surveys on diverse subjects. Usually they serve as the first stage in three-stage selections of districts x ED's (enumeration districts) x segments. Local interviewers haved been hired, trained and employed as needed. The problem is that a few districts are close to being "exhausted" : districts where most segments in most ED's have already been used in surveys, and we don't want to go back to them again. However, we don't want to discard now our sample of districts, because most of them have more than enough ED's until we select a new sample in 1982. Furthermore, we believe that to simply make another selection with PPS in the strata of the "exhausted" units would be biased against small districts. Is there an unbiased procedure that would allow us to substitute for the few exhausted districts, yet save the great majority of PSU's which are far from exhausted and which are of great value for us ?

Answer by Leslie Kish, ISR, U. of Michigan, Ann Arbor, Mich. 48106, USA

Yes, there is. And you note correctly that it is a biased procedure to simply substitute with another PPS selection for those selections which turned out to be smaller than sufficient. That is a fairly common mistake that can and should be avoided.

It is useful to link the problem of "exhausted" units with the more general problem of "undersized" units which lack "sufficient size", a problem that often occurs in PPS selection. That link is useful because the same solution can help us avoid both problems with foresight at selection time, or alternatively to repair them with hindsight, if need be. The link also shows that in many situations it is preferable to plan ahead with a large sufficient size than needed immediately, because the same selected PSU's may be wanted later for other surveys as well.

The problem is of greatest importance for selections of PSU's (like districts), but can occur also with secondary units (like blocks). The PSU's are selected with probabilities proportional to the measures of size Mos_i , then these are subsampled with rates of b/Mos_i to yield an equal overall selection probability of $1/F$ in two stages from the stratum :

$$\frac{Mos_i}{\sum Mos_i} \cdot \frac{b}{Mos_i} = \frac{b}{\sum Mos_i} = \frac{1}{F}$$

Here b represents the desired "sufficient" size, after considering factors of costs and variances. If a PSU were selected with $Mos_i < b$, it would result in a subsampling rate of $b/Mos_i > 1$, which is not feasible. Then the sampler after taking the entire cluster would be faced with the unpleasant choice of either accepting a bias of factor Mos_i/b, or of restoring an unbiased expectation with that factor.

That unpleasant choice can be avoided by using a minimal "sufficient" size b. This can be done in the initial selection with two procedures [Kish, 1965, Section 7. 5E on undersized units] : (a) Create separate strata for units with $Mos_i \leq b$ (or somewhat larger for convenience), and select them with $1/F$ for sampling complete (compact) units ; or (b) assign arbitrary measures $Mos_i = b$ to all units with $Mos_i < b$ (or even slightly larger) for complete subsampling.

In the last two paragraphs the sufficient size b was set to yield some specific value for one or more samples which would exhaust the units with $Mos_i \leq b$. But b can be set more generously and that will permit expanding its future use, also yield more flexibility in general. It results in a lower subsampling rate b/Mos_i to avoid high subsampling rates which may be uncomfortable in some situations.

What do we pay for the greater comfort of a larger minimal b ? That depends on the situation, but let us suppose a typical situation of selecting districts with PPS. The measures are highly unequal, with the Mos_i for most selected units much larger than b, some by factors of 100 or more. Only about 3 of 120 selected units would have $Mos_i < b$. These selections with PPS represent the distribution of the population elements, although only a small portion of the units are large and many are small. Doubling the minimal size to 2b may increase the number of selected units with minimal size to only 5-7 or so, and would not greatly increase the costs of subsampling.

The two procedures above (separate strata and arbitrary measures) can be used to plan ahead for sufficient sized units, but not readily to correct later for exhausted units. However linking procedures may be used either to create units with sufficient sizes before selection, or to correct later for exhausted units with similar procedures.

Linking procedures may be used to create new linked units (PSU's) with $Mos_i \geq b$, which combine arbitrarily units with $Mos_i < b$ with other units. They may be linked in several ways with different aims.

1) Adjoining units may be linked to form convenient units so that the same field workers can handle readily the linked units. These linkings may conflict with the stratification of the units, and this problem must be handled before selection.
2) The listing order of units may be used for linking units conveniently to form units of sufficient size.
3) Similar size units may be linked and regarded as linked units.

When the linking is done for sufficient size before selecting the linked units then one may be arbitrary about which units are linked. But if linking must be done after selection, as for exhausted units, then we need objective procedures which avoid bias. Objectivity can be had either with ignorance or with strict procedures, either of which excludes the past selection from influencing the linking. Thus : (1) Linking adjoining units can be done by somebody, otherwise well trained for the task, who is, and is kept, ignorant of the names and locations of the selected units in the stratum. (2) The listing order may be used to link with the last sufficient unit any and all insufficient units up to (but not including) the next sufficient unit [Kish, 1965, 7.5E]. (3) The selected unit can be linked to the next unit (in a circular order, the first follows the last) which has a similar size (as defined beforehand).

After units linked for sufficient size have been selected they can be either (a) subsampled as a single PSU, or (b) unlinked, one unit selected with PPS then this subsampled. The latter (b) is the needed model for exhausted units, because its exhaustion is equivalent to it. Now the newly linked unit(s) must be regarded as selected, and subsampled accordingly.

For example, suppose you now want $b_{min} = 200$ as the minimal sufficient size ; that you had exhausted a district with $Mos_i = 120$; also that a linking procedure gives a linked unit of three districts with its $Mos_i + 120 + 60 + 50 = 230$. Now the last unit can be selected with probability $50/(60 + 50)$. Thus :

$$\left(\frac{230}{\sum Mos_i} \times \frac{110}{230} \times \frac{50}{110}\right) \times \frac{b}{50} = \frac{b}{\sum Mos_i} = \frac{1}{F}$$ describes the overall selection probability, with the first three terms for selecting the linked unit, the non-exhausted districts, and the single district. The subsampling rate $b/50$ is the same we would have had with separate selection of the district, or with a foresighted linking followed by a selection with $50/230$ of the district.

For simplicity we assumed single selections from each half-stratum, and this was true for the sample in the question. (Pairs of half-strata are used for variance computations). The procedure would be similar for selections of two units with PPS with replacement. But for complex procedures of two or more selections without replacements the linking procedures could become difficult. (Complex procedures often cause unforeseen problems later, alas).

Exhausted units are one of the problems involved with continuing sampling operations [Kish, 1965, 12.6-12.7]. Related to this question is another which concerns retaining units with changed strata and probabilities. When the new sample is selected in 1982, it is not necessary to select an entirely new sample, and it may not be desirable. This problem was raised in diverse forms in the Workshop and will be discussed in an early issue [Kish and Scott, JASA, 1971.].

References :

Hansen, M.H., Hurwitz, W.N. and Madow, W.G. [1953], Sample Survey Methods and Theory, Vol. I, New York : John Wiley and Sons, p. 347.
Keyfitz, N. [1951], "Sampling with Probability Proportional to Size ; adjustment for Changes in Probabilities", JASA, 46, 105-109.
Kish, L. [1965], Survey Sampling, New York : John Wiley and Sons, Sections 12.6-12.7.
Kish, L., and Scott, A. [1971], "Retaining Units after Changing Strata and Probabilities," JASA, 66, 461-470.

NEWS FROM THE ASSOCIATION

JUNE 27 1978 MEETING IN PARIS

On June 27, 1978, a meeting took place at INSEE in Paris to which participated : Dr Sukhatme, IASS President, Dr Sanchez-Crespo, Chairman of the Nominating Committee, Mr Théodore, Chairman of the Publication Committee, Mr Damiani, Administrative Secretary.

During the meeting, the following issues were discussed :

- pattern and contents of the new publication "The Survey Statistician" that will replace the Newsletter
- list of candidates to the Bureau and Council for the 1979 election, presented by Dr Sanchez-Crespo
- financial situation and number of members in the Association
- results of the vote on the amendments to the Statutes presented at the New Delhi Session in 1977 : the amendments were approved
- preparation of an interview between Dr Sukhatme and Mr Théodore with a representative of the non governmental organisms office of UNESCO for an application to UNESCO from IASS
- nomination of an Executive Director to replace Dr Narain who resigned : Dr Sukhatme proposes the post to Mr Théodore who accepts but only in the interim up to the next Session in Manila in 1979.

OBITUARY

The Bureau of the Association regrets to announce the death of Dr Tekse Kalman, Director of the Central Institute of Statistical Research, Candidate of Demographic Sciences in Hungary who died on August 3, 1978.

BIBLIOGRAPHY

Organization for Economic Co-operation and Development. -

"Emploi, durée du travail et salaires"- Main Economic Indicators, Sources and Methods, n° 29, May 1978, 116 p.

Food and Agriculture Organization of the United Nations . -

"National Methods of Collecting Agricultural Statistics"- Asia and the Far East Commission on Agricultural Statistics, Bangkok, Thailand, 17-23 August 1978, 380 pages.

United Nations Economic Commission for Western Asia . -

"Population Bulletin" - n°13, July 1977, 68 pages :
- Religious Fertility Differentials : a Look at some Hypotheses and findings, Joseph Chamie
- Female Education and Fertility Decline in a Developing Country : the Case of Jordan, Afaf Deeb Kandis
- The Effect of Education of Women and Urbanization on Actual and Desired Fertility and on Fertility control in Lebanon, Huda C. Zurayk
- The Demography of the Population of Kuwait, Allan G. Hill
- Census Data Required for Indirect Methods of Estimating Demographic Parameters - 1980 Round of Censuses, Ken Hill
- Book Reviews and Periodical Abstracts

To get copies, request may be addressed to : the Population Division of ECWA, P.O. Box 4656, Beirut, Lebanon

Philippe L'Hardy . -

"Erreurs de mesure dans l'évaluation du taux d'épargne par catégorie de ménage" - CREP-INSEE International Meeting, Wealth Accumulation and Distribution, Paris July 5-7, 1978, Direction des Synthèses Economiques, Unité de recherche, INSEE, 65 pages.

M.H. Williams . -

"How bad can good data really be ?" - The American Statistician, May 1978, Vol. 32, n°2, p. 61-65.

P. Thionet . -

"Quelques problèmes concernant les sondages" - Vanderhoeck & Ruprecht, Göttingen (R.F.A.), 1978, 137 pages.
What is the origin of a theory of statistics, the mathematical sampling survey one ? How it first used the existing theories relating to other problems (men too) ; then how new branches appeared progressively, the motivations being theoretical more than practical.
This science is owed to the Anglo-Saxons, the Indians and the Scandinavians, - and some French too : the real topic of the book is to replace the small French group into the international context ; during 25 years they dealt with all the topics ; their efforts seem now completed.
Let us point out that the publication "Etudes Econométriques" (Lyons) will issue soon, by the same author a review of the internatioal bibliography of mathematical sampling surveys during the 1967-77 period, - a work completing the present book actually.

K.J. Krotki, S.S. Hashmi . -

"Issues in Demographic Data Collection in Pakistan"- Proceedings of a Seminar , Rawalpindi, 11-13 August 1975, XXXV + 342 pages, 53 tables, 3 figures, 26 appendices, hardcover.
The 24 authors, predominantly Pakistanis, are demographers, geographers, cartographers, family planning analysts, analytic statisticians, survey statisticians, and administrators.
Among the topics covered are :
Vital events registration under the National Registration Act, 1973 ; Survey as a permanent data source of Pakistan's vital events ; Dual estimation as a permanent data source of Pakistan's vital events ; the Census Evaluation Survey of 1972 in the light of experience with similar endeavours elsewhere ; Age recording for the same person at different times in a Punjab village ; Impact of age misreporting on fertility estimation ; Application of own children technique for fertility estimation in the HED survey ; Pakistan participation in the World Fertility Survey ; Population registration as a demographic data source ; Conceptualization of urban areas for the 1981 census ; Information feedback system for population planning in Pakistan ; The past and future of Pakistan's demographic data.

For copies apply to : Mr G. Mujtada Mirza, Director, Census Organization, 16 Almarkaz Building, Islamabad, Pakistan

Statistics Canada . -

" Survey Methodology" - Statistical Services Field, Vol. 3, n°2, December 1977
- Confidentiality of Statistical Information, D.A. Worton
- Synthetic Estimation in Periodic Household Surveys, P.D. Ghangurde, M.P. Singh

- L'enquête sur la profession des salariés (E.P.S.), M.A. Hidiroglou
- Some Factors Affecting Non-Response, R. Platek
- Survey of Canadian Residents Returning byland, J.H. Gough, P.D. Ghangurde
- An Investigation of the Properties of Raking Ratio Estimators II With Cluster Sampling, H.R. Arora, G.H. Brackstone

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

The International Association of Survey Statisticians (IASS) was created in 1973. The IASS is a section of the International Statistical Institute.

The objectives of the Association shall be to promote the study and development of the theory and practice of statistical censuses and surveys and associated subjects and to foster interest in theses subjects among statisticians, organizations, institutions, governments, and the general public in different countries of the world.

OFFICERS AND MEMBERS OF THE COUNCIL

President (1978-79)	P.V. Sukhatme (India)
President elect (1978-79)	T. Dalenius (Sweden)
Executive Director (acting) (1978-79)	G. Théodore (France)
Vice-presidents (1978-79)	I. Fellegi (Canada)
	L. Kish (USA)
	A.J. Scott (New Zealand)
Scientific Secretary (1978-79)	G. Théodore (France)
Treasurer (1978-79)	E. Lunenberg (Netherlands)
Director of the ISI Permanent Office	
Members (1978-79)	E. de Alba (Mexico)
	F. Azorin (Spain)
	S.L. Diop (Senegal)
	J. Harewood (Trinidad)
	S.H. Khamis (Lebanon)
	M.N. Murthy (India)
(1978-81)	W.E. Deming (USA)
	C. Hayashi (Japan)
	D. Holt (United Kingdom)
	K.J. Krotki (Canada)
	G. Nathan (Israel)
	A.S. Nielsen (Denmark)

SECRETARIAT

Mr Paul DAMIANI, IASS, c/o INSEE, 18 boulevard Adolphe Pinard, 75675 PARIS CEDEX 14, France

THE "SURVEY STATISTICIAN"

The Survey Statistician has replaced the IASS Newsletter. It is published periodically by INSEE, in English and French

The Survey Statistician and Sankhya C represent the IASS official publications. While Sankhya C contains theoretical papers, the Survey Statistician includes practical information in the field of sampling surveys (surveys in progress, summaries of technical and methodological papers, answers to questions, bibliography).

Editor : J. Desabie, Inspecteur général de l'INSEE, INSEE, 18 boulevard A. Pinard, 75675 PARIS CEDEX 14, France

Editing Board : Messrs Dalenius, Desabie, Kish, Théodore

INFORMATION FOR PAPERS' AUTHORS

The papers sent to the Journal should not exceed 5 pages. They have to be typed with a double space. Two copies of these papers (in French or English) have to be sent to Mr Damiani, at the IASS Secretariat (see address above).