

**INTERNATIONAL STATISTICAL INSTITUTE
INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS**

SURVEY STATISTICIAN N° 1

FEBRUARY 1979

| SUMMARY | Page |
|--|------|
| <u>Editor's letter</u> | 2 |
| <u>General information</u> | 2 |
| - Program of the IASS meetings for the 1979 Manila Session | |
| - IASS Workshops | |
| 1. Activities of the IASS Workshop Committee | |
| 2. IASS Regional Workshop in Survey Sampling, Mexico, Nov. 6-10, 1978 | |
| 3. IASS Regional Workshop in Survey Sampling for Asia | |
| - UN expert meeting on integration methods of demographic and social statistics, March 27-31, 1978 | |
| - 3rd Session of the Statistical Committee of the Economic and Social Commission for Asia and the Pacific (ESCAP), Bangkok, October 17-23, 1978 | |
| - 8th Conference of Statisticians of the Customs and Economic Union of Central Africa, Brazzaville, October 16-21, 1978 | |
| <u>Surveys in Progress</u> | 6 |
| - Surveys in Progress in the U.S.A. | |
| • 1974 Point of Purchase Survey | |
| • 1972-73 Consumer Expenditure Surveys | |
| • Consumer Price-Index Revision - Rent | |
| - Study on Incomplete Data, U.S.A. | |
| - Surveying the Extent of Drug Use, Australia | |
| <u>Training</u> | 12 |
| Theoretical Preparation and Methods of Selection of Survey Statisticians : Spanish Experience - J.L. Sanchez-Crespo | |
| <u>Terminology</u> | 14 |
| Populations for Survey Sampling - L. Kish | |
| <u>Papers</u> | 15 |
| - The Swedish Party Preference Surveys - S. Sollander | |
| - Household Surveys in France : on the 1976 master-sample - F. Chartier | |
| - Correlation Ratio and Sampling Design - M. Deroo | |
| <u>Question/Answer (Editor : L. Kish)</u> | 22 |
| What are the methods which have been used in the enumeration of nomads or semi-nomads in censuses or surveys in Africa and what are their advantages and disadvantages ? Answer by K.T. de Graft-Johnson | |
| <u>News from the Association</u> | 24 |
| - Visit of personalities in Paris | |
| - Preparation of the program of the meetings of the 1981 ISI-IASS Session | |
| <u>Bibliography</u> | 24 |

EDITOR'S LETTER

The Survey Statistician replaces the IASS Newsletter edited since 1974. The Newsletter n°8, edited in October 1978, was the last issue ; it was called too "Survey Statistician n°0" since it was a tentative issue for the latter, concerning the presentation and the presented sections.

The Survey Statistician is a periodical publication published twice or three times a year, according to the number of papers and news to be edited.

The Editing Board would be glad to receive comments from IASS Members in order to improve the presentation of the publication together with the quality of the sections. He reminds that this newsletter cannot live without the contribution of papers from its members.

GENERAL INFORMATION

PROGRAM OF THE IASS MEETINGS FOR THE 1979 MANILA SESSION

The overall program for the Manila session calls for 30 scientific meetings. Below is a list of the meetings organized by the IASS :

| <u>Titles</u> | <u>Organizers</u> |
|---|---------------------------------------|
| a. Technical information exchange on statistical software (TIESS) with emphasis on survey and census processing | I. Francis (New Zealand) |
| b. Interactive computing, interactive procedures with large data files | I. Francis (New Zealand) |
| c. Rotation and other resampling schemes | I.P. David (Philippines) |
| d. Incomplete data | W.G. Madow (U.S.A.) |
| e. Estimation of population growth | W. Brass (U.K.) and G. Calot (France) |
| f. Enterprise statistics | G.K.G. Forbrig (G.D.R.) |
| g. Privacy and confidentiality issues in surveys | T.J. Jabine (U.S.A.) |
| h. Market intelligence surveys | S. Tulya-Muhika (Uganda) |
| i. Surveys in developing countries | M.N. Murthy (India) |

Meetings a and b are, by confirmed agreements, organized jointly by the IASS and the IASC (International Association of Statistical Computing).

Meetings d and e are expected to be organized jointly by the IASS and the ISI.

IASS WORKSHOPS

1. Activities of the IASS Workshop Committee

L. Kish (President), I.P. Fellegi, M.N. Murthy, R.D. Narain, H. Nisselson

One workshop in Spanish for the Ibero-Hispanic region was held in Mexico. Another for the Asian region is being planned in English in Honolulu. Note for each that :

- Advantage is being taken of other opportunities with resources to attach IASS workshops to them. We need to do this until resources are found for separate financing (if ever).
- Geographical proximities are combined with linguistic considerations to arrive at informal regional groupings.

We suggest that similar attempts could be made in four areas in Africa : French West, English West, English East, and an Arabic Northeast and North, together with southwest Asia. In Europe it would be possible to have four major groupings : Slavic, Scandinavian, Germanic, and French. These are informal and personal suggestions, without diplomatic, formal, or permanent meanings.

from L. Kish

2. IASS Regional Workshop in Survey Sampling, Mexico, November 6-10, 1978

The principal aims of the meeting were improving regional contacts by means of discussions of practical problems in survey sampling raised by participants from eight countries. There were nine participants from Brazil, Chile, Colombia, Ecuador, Peru, Spain, and the United States, in addition to about 50 from Mexico itself.

In addition, there were addresses by Francisco Azorin and Jose Sanchez-Crespo from Spain ; Enrique Cansado, Chile ; and Leslie Kish, U.S.A.

A Committee on Survey Sampling in Iberico-America (for Latin America plus Spain plus Portugal, as a language grouping) was formed to promote more regional meetings, to establish a regional directory of survey statisticians, to spread in Spanish recent developments in survey methods, and to support established organizations : IASS, IASI, and COLPRIE. As temporary members to head the Committee, the assembly named : Edmundo Berumen, Mexico (Chairman) ; Francisco Azorin, Spain ; Eunice Pinho de Castro, Brazil ; one more from Venezuela to be appointed soon. The committee represents intentions to hold the next two annual regional meetings in Brazil and in Venezuela.

For future workshops please note that membership in IASS was solicited and applications were sent to the Paris office.

For visitors from abroad, allowances for hotel and meals were provided by an agency in the host country. Travel was provided by offices of the participants. It was noted that visitors from abroad were attracted through personal solicitation, whereas notices through official channels to the statistical offices proved entirely fruitless this time, alas.

from L. Kish

During the Mexico Workshop, among others the following papers were presented : "Algunas experiencias de muestreo en el Instituto Nacional de Estadística de España" (Some sampling experiences in the Spanish I.N.E.)" by F. Azorin and J.L. Sanchez-Crespo. The content of this paper is as follows : continuous surveys ; the GPS, the frame, organization of the field work, type of sampling, sample size and rotation groups, sampling errors. Processes of stratification, selection and estimation, small area problem, updating of the frame and the probabilities of selection. Programmes of evaluation in censuses and surveys ; pilot census, sample design in the evaluation of the census ; evaluation of the GPS : sampling errors, non sampling errors : coverage and content errors.

"Estimación, con probabilidad gradual, en poblaciones finitas" (Estimation with gradual probability in finite populations)" by J.L. Sanchez-Crespo. The paper deals with : unequal probability sampling, with replacement, without replacement, and with probabilities gradually variable ; gain in precision. General formulae which contains linear and non linear estimators.

"Muestreo de Poblaciones Borrosas" (Sampling in fuzzy populations)" by F. Azorin, with the following content : Introduction : concepts, definitions, notations ; fuzzy populations and fuzzy samples ; fuzzy clustering. Fuzzy information. Fuzzy inference and fuzzy decisions : desiderata and evitanda, fuzzy population values and fuzzy estimators, sampling and non-sampling errors for fuzzy populations. Applications, trade-offs and conclusion.

"Muestras y censos" y "El problema de las areas pequenas" ("Samples and censuses" and "The problem of small areas") by L. Kish. In the first paper, the author presents a battery of criteria to consider the advantages and inconvenients of different sources of data, such as censuses, sampling surveys and registers. In connection with the second paper, several procedures of estimation are presented and discussed.

About 50 participants from Latin-American countries were also invited and several interesting problems were presented for discussion.

It was decided to create the "Comite Iberoamericano de Muestristas" (Ibero-American Survey Samplers Committee) which will include samplers from Latin-America, Portugal and Spain. The committee would promote sampling techniques in Spanish and Portuguese speaking countries.

The proposal that Professor Kish should be named Honorary President of the Committee was met with widespread satisfaction.

from J.L. Sanchez-Crespo

3. IASS Regional Workshop in Survey Sampling for Asia

At the East-West Population Institute, there will be held January 24-February 16, 1979, a Census Sampling Workshop. The workshop will focus on practical aspects of sampling designs and problems in connection with the 1980 Censuses in diverse countries. The session will be coordinated by Leslie Kish from the USA and Jae-Soo Park from Indonesia. For information write to Census Sampling Workshop, East-West Population Institute, 1777 East-West Road, Honolulu, Hawaii, 96845, USA.

Participants are expected from 12 or more Asian countries. We want to take advantage of this opportunity to hold an IASS Workshop on practical problems of survey sampling with the additional aims of increasing contacts on a regional basis among survey samplers. We shall set aside three days, February 5-7, for this purpose. We can make a modest effort to invite a few participants in addition to the above, but no funds are available for the purpose. Those interested should write to the above address.

from L. Kish

U.N. EXPERT MEETING ON INTEGRATION METHODS OF DEMOGRAPHIC AND SOCIAL STATISTICS, MARCH 27-31, 1978

The main topic of this meeting of experts organized by the U.N. on March 27-31, 1978, was the review of file merging methods.

The Members of the group were experts from : U.S.A. (1), Canada (2), Brazil (1), Kenya (1), Sweden (1), Denmark (1), Poland (2), France (1). Some observers, the representatives of specialized organizations of the U.N. took also part in the meetings.

A "preparatory document" was used as a guide to the discussions ; a very interesting documentation on the experience of participating countries in the field of file merging was annexed (a very brief summary of it is given in fine).

1. EXACT MATCHING AND STATISTICAL MATCHING

Concerning file merging, a distinction is made between "exact matching" realized thanks to a common id-number and "statistical matching".

The advantage of linking individual data thanks to an id-number is obvious to any statistician. However, it was clearly stressed out that any matching, even an exact one, includes a part of randomness. Confidentiality problems were evoked in quite a large (but dim) way.

File "statistical matchings" are a topic of larger intellectual interest. The principle of the method is as follows : given two sources providing, at an individual level :
one, the data A and the data B
the other, the data A and the data C,
the statistician takes responsibility to postulate that for two individuals close with regard to A (as a matter of fact, the A vector will generally include two components : the A₁ one, for which identity is demanded, the A₂ other one, for which proximity will be enough), there is a statistical independence between B and C ; in another way, he decides that the matching of the individuals of two files belonging to the same stratum A can be done at random. Thus, through random matching we realize individuals for which A, B and C are "known simultaneously".

The similarity with the intellectual processes followed in the field of weighted estimator (1) sample is obvious : similar independence hypothesis :

in one case, between the attitude toward the survey (acceptance or refusal) and the observed data,
in the other, between two sets B and C of observed data,

rejected for the population in its whole, but postulated, under the statistician's responsibility, for the individuals of a same stratum A.

However, an important difference is that the sample weight only introduces a margin correction for the favourably welcomed surveys. On the other hand, the objective, much more ambitious, of file mergings is, actually, to create an extra information by supposing gathered at the individual level, the information B and C that actually could not be obtained simultaneously at that level because :

- either the surveys they are coming from are provided by independent samples
- or obstacles of deontological nature appear against the merging from a common id-number of two files containing the wished information for the same individuals.

The interest of the method was fervently demonstrated : thus it would be possible to create data banks that would allow to meet, at any price, because that is the point, the demand ... while yielding to the pressures of "privacy" defenders (2).

2. QUESTIONS ARISEN

However this "statistical matching" solution remains questionable for some reasons.

Indeed, either the independence hypothesis between B and C for the individuals of a same stratum A is founded, then what is the interest of merging files or of carrying out a survey giving A, B and C simultaneously ? Indeed, for a given A, we already know the first two moments of distribution of B for the file AB, the first two moments of distribution of C for the file AC and we think that B and C are independant. Thus, we already dispose of all the desired information. The merging does not yield any new information but only calculation facilities that besides can be very profitable.

Or this independence hypothesis is wrong, in which case instructive deductions made from the merging are wrong.

Such a risk is not imaginary, specially because of "budget constraints", in a wide sense.

Whatever the nature of the constraint : income, time, available spaces, implicit nutritional standards, it is opposed to the independence between B and C for any stratification practically realizable.

In a larger sense, if between the inputs and outputs respectively appearing between B and C, there are complementarity or substitutability relations, the bias risk is obvious.

Thus, let us suppose - the example is realistic - that we mean to merge specialized surveys dedicated, each, to a savings component, we would meet the difficulty due to the facts that :

- some investments are quite competitive, aiming at a same objective, in somewhat different ways

(1) taking into account non responses.

(2) This type of secret, respectable, is not threatened in the least, since statisticians' works are nearly always performed on samples.

- while others are closely complementary such as in the case of an intermediary saving caused by the refunding of a real estate loan.

Another example, quite different, given by one of the representatives of the U.N. Statistical Office, to persuade us of the interest of the method, is quite convincing too. Let two tables be :
AB : professional income x socio-professional individual category and statute,
AC : professional income x training.

Taking into account the practices of the French Administration, the independence hypothesis is surely wrong in our country.

It is possible to get an interesting table by merging.

If we point out that the coding of A may be somewhat heterogeneous in the two sources, we see that the risks of illegitimate mergings are real.

At that level, the necessity to test the independence hypothesis was then outlined. This can only be done by reduced surveys giving A, B and C simultaneously. It is then possible to split the file at random into two files playing the part of the file AB (C is ignored) and the file AC (B is ignored) to simulate the merging and compare the result thus obtained (concerning the table BC) to the "true aspect" given by the original file ABC.

There too, the decisive advantages of the observation should never be forgotten by the statistician.

from J. DESABIE
Head of the Department "Population-ménages"
INSEE

THIRD SESSION OF THE STATISTICAL COMMITTEE OF THE ECONOMIC AND SOCIAL COMMISSION FOR ASIA AND THE PACIFIC (ESCAP)
BANGKOK, OCTOBER 17-23, 1978

The Committee heard the General Secretary's Report mainly dealing with the following issues :

- organization of the 1980 population and dwelling censuses
- regional workshops on the computer tabulation of censuses and surveys (ISH)
- seminar on rural development statistics
- seminar on small and handicraft industry statistics (July 1978)
- document and file compilation and maintenance for development projects.

Besides, the Committee dealt with the following topics :

- training needs and means in the region
- project of revision of the standard international classification by industry
- setting up of national means for household surveys
- statistics necessary to the development strategies for the 1980's and the middle term plan (1980-83).

EIGHTH CONFERENCE OF STATISTICIANS OF THE CUSTOMS AND ECONOMIC UNION OF CENTRAL AFRICA, BRAZZAVILLE,
OCTOBER 16-21, 1978

Appeared on the agenda the following issues :

- report of activities and program of the Statistical Department of the General Secretariat
- organization of the National Statistical Services
- training of Statisticians
- regional and sub-regional statistics
- improvement in current statistics and the circulation of statistical information
- health statistics
- education statistics
- national files of villages
- birth and death statistics
- regional project of agricultural census
- classification of economic agents
- short statistical declaration of enterprises under the contract tax system
- firm files
- budget-consumption expenditure surveys.

Further to the discussions, the Conference adopted six recommendations concerning :

- the training of statisticians
- national file of villages
- classification of economic agents
- short statistical declaration
- registration of enterprises and establishments.

SURVEYS IN PROGRESS

SURVEYS IN PROGRESS IN THE U.S.

1974 Point of Purchase Survey (U.S.A.)

In the early 1970's the Bureau of Labor Statistics in cooperation with the Census Bureau conducted a series of household surveys. The 1974 Point of Purchase Survey (POPS) along with the 1972-73 Quarterly and Diary Expenditure Survey were used by BLS for the revision of the Consumer Price Index (CPI) in 1978.

The Point of Purchase Survey (POPS) was designed to provide the sampling frame of outlets for food and for most commodities and services to be priced in the CPI. Formerly, existing directories of retail outlets were utilized to select a sample of outlets for pricing.

The POPS Survey was conducted in an 85 Primary Sampling Unit (PSU) area design which corresponds to the Consumer Price Index PSU design. This design resulted in 27 self-representing strata and 58 non self-representing strata. In selecting the one sample PSU for each non-self-representing stratum, a controlled selection program was used to insure that the sample areas were properly distributed geographically across states. Since separate frames of outlets were required for certain individual pricing areas (PSU's), sample sizes were not proportional to population. The POPS "national" household survey is not self-weighting, however, within a PSU the household sample is selected with a uniform probability.

When determining which POPS categories were to be inquired about on the point of purchase questionnaire, special emphasis was given to apparel, food, services purchased frequently and common commodities.

The total sample size of the (POPS) Survey was 23,000 interviewed households. The number of households selected for PSU's was chosen such that it yielded an expected number of outlets responses for a POPS category equal to three times the desired number of expected outlets to be selected.

With the introduction of the household survey approach, it was decided to highly cluster sample households. If families buy in areas where they live, the outlets given as responses would also be clustered. This was expected to reduce CPI data collection costs.

A PSU was divided into SSU's (secondary sampling units), an SSU being a set of contiguous census tracts around a shopping complex. These shopping complexes were Central Business Districts, Major Retail centers (as defined by the 1967 Census of Business) and other shopping centers (as defined by the 1970 Directory of Shopping Centers).

Each SSU was assigned an economic measure which was a weighted average of the median family income of the tracts in the SSU as reported in the 1970 census. The weights were 1970 Census housing unit counts. SSU's were then ordered based on median family income.

A sample of SSU's was selected with probability proportional to size to the 1970 housing unit count of the SSU. This economic-geographic stratification was undertaken in order that the sample of SSU's would be a better cross-representation of the entire PSU.

Clusters of households were formed around shopping complexes. Within a cluster of census tracts, a sample of EC's was selected and within the selected ED's (Census enumeration districts), the sampled households were systematically selected.

To represent units constructed since 1970 the Census Bureau sampled new construction off register permits without regard to SSU. Sample new construction units were spread throughout the PSU.

The total cost of the POPS Survey was \$ 1.8 million dollars. All micro information is confidential. The survey supported the all urban index for the CPIR as designed, but provided an inadequate sampling frame for the urban wage clerical CPIR family definition in 5 % of the categories and for which it was not designed. BLS is continuing its use of a POPS like survey to obtain a sampling frame for outlets for the CPI with modifications of sample size and reference periods for expenditures for outlets to support the urban wage clerical population. The only POPS category which totally failed to support an adequate frame was maintenance and repair of homes. A significant proportion of non-professional part-time private individuals were reported and are subsequently out of scope for continued repricing.

Further information concerning the Point-of-Purchase Survey may be obtained by writing to Miss Barbara A. Boyes, GAO Building, 441 G. St., N.W., Washington D.C. 20212 or by telephoning 202.523.1098.

1972-73 Consumer Expenditure Surveys (U.S.A.)

The 1972-73 Diary and Quarterly Consumer Expenditure Surveys were conducted by both the Bureau of Census and the Bureau of Labor Statistics for the purpose of updating the data base on expenditure patterns as reflected in the Consumer Price Index.

The Diary collected data on expenditures for items purchased on a daily or weekly basis. Frequently purchased items included food for consumption at home, food for consumption away from home and personal care expenditures. To accomplish this, a sample of households was asked to maintain records on all purchases made by week, for a two week period with the sample of households being spread evenly throughout the years of 1972-73.

The Quarterly collected data on expenditures for items purchased less frequently than on a daily or weekly basis. Examples of such items are transportation, medical expenses and household expenses. For this survey, three panels of sample households were established to spread data collection throughout each quarter for five consecutive quarters. The recall period for reporting expenditures was set by the size of frequency of the expenditure class. For the more frequently purchases and less expensive items, data was collected more often than the less frequently purchases and more expensive items.

The eligible population was composed of all civilian noninstitutional persons and certain institutional persons. Only institutional persons living in rooming and boarding houses or in nurses' and doctors' quarters of general hospitals were included in the eligible population.

The nation was divided into 216 geographic strata. Thirty Standard Metropolitan Statistical Areas (SMSA's) with a 1970 population of 660,000 or more were self-representing strata. Less populated areas were grouped into 186 strata. From each of the 186 strata one primary sampling unit (PSU) was randomly selected using a controlled selection procedure to insure proper geographic distribution.

The designated sample of housing units selected for the 1972 and 1973 Diary were 14,586 and 15,707 respectively. At the time of selection, housing units within a PSU were distributed by two ~~week-periods~~ for interviews throughout the year. Each week, each consumer unit (CU) within the household compiled a diary of expenditures. A consumer unit is a consumer or family of two or more persons living together, pooling income and drawing from a common fund for major expenditures.

For the Quarterly, a sample of 12,613 housing units was designated for the 1972 Quarterly and 13,014 housing units for the 1973 Quarterly. At the time of selection, housing units within a PSU were distributed by month within the quarter to allow for data collection throughout each quarter. Each sample unit was visited once each quarter and each consumer unit within the household was interviewed.

Each eligible housing unit and institutional person was assigned to a stratum based on the 1970 Census of households. Occupied housing units were stratified by income level, tenure, and family size. Vacant units which were for sale or rent were assigned to separate strata, and those vacants which were not for sale or for rent were assigned to a different stratum. Eligible institutional persons comprised the final stratum. Sampling was performed independently within each PSU, and the selection of institutional persons was independent of the selection of housing units. In addition, a number of newly constructed housing units were selected from building permits to update the sample for the period from 1970 Census to the time of interview for the survey. Again, the sampling was performed independently within the PSU.

The samples which had been selected for the Diary and Quarterly were divided into two parts. Half of the sample units for the Diary in each of the thirty self-representing PSU's were covered in the period from July 1972 to June 1973 (the 1972 Diary) and the other half of the units were covered in the period from July 1973 to June 1974 (the 1973 Diary). For the Quarterly, half of the sample units in the certainty PSU's were covered from Jan. 1972 to Mar. 1973 (the 1972 Quarterly) and the other half of the units were covered in the period Jan. 1973 to Mar. 1974 (the 1973 Quarterly). The remaining 186 non-self-representing sample areas were divided into two panels, each with 93 areas. One panel was covered in each phase of the two surveys.

A weight was determined for each person within a consumer unit (CU). This weight depended upon such factors as the inverse probability of selecting a housing unit, whether the person was interviewed in the pre-Christmas holiday month for which a special supplement sample was selected, if the CU was affected by subsampling in the field and several ratio estimates. The final weight for residence, age, sex, and color of an individual CU was determined by the type of family composition of the CU.

The combined cost of the surveys was \$ 1.8 million and was a significant improvement over previous expenditure surveys of this type. The only significant problem was the integration of the two surveys for expenditure by income tabulations since income was not reported in the same manner.

Data tabulations are available upon request from BLS as well as a public use tape.

Further information concerning the 1972-73 Consumer Expenditure Surveys may be obtained by writing to Miss Barbara A. Boyes, GAO Building, 441 G. St., N.W., Washington D.C. 20212 or by telephoning 202-523-1098.

A. Consumer Price Index Revision (U.S.A.)

The Consumer Price Index has undergone an extensive revision. Plans for the revision began in 1968, a new Consumer Expenditure Survey began in 1972, the POPS Survey for selecting outlets to be priced began in 1974 and the publication of the first index from CPFR Revision was in February of 1978. Periodic revisions of the index are necessitated by several factors. Purchasing habits change as tastes change. New products are introduced and old ones become obsolete. Buying patterns also changed because prices for different items change at different rates. The outlets patronized change as well as the populations shift and new types of stores appear.

This paper outlines the sampling procedures to select the items, outlets and specific store items within outlets for the two family indexes.

The expenditure data for the CPI is constructed from the Diary and Quarterly Consumer Expenditure Survey conducted by Census for BLS in 1972-1973. The basic structure created for both the Diary and the Quarterly was to define 68 Expenditure Classes (EC's) which are BLS's primary definition of publication levels of indexes. Each EC contained one or more item stratum and within each item stratum, one or more substrata called Entry Level Items (ELI's). The ELI's are the sampling unit used by the data collectors as their initial level of aggregate item definition within an outlet. ELI's are broadly defined groupings of items and allow the possibility of pricing many different kinds of items within a store.

This structure of the Diary and Quarterly into item strata and ELI's is the same for all regions populations and publication areas ; however, for sampling purposes, BLS tabulated four regional universe market baskets for each population to reflect regional differences. Within each region, BLS made eight independant selections of ELI's within each item stratum forming eight "half-samples" of ELI's. These half-samples were distributed among the CPI's 85 pricing areas (PSU's) for pricing within the region.

Because of the requirement of publishing two family indexes each selection of ELI's within an item stratum was made initially for the urban wage and clerical population (W) proportional to the relative expenditures of ELI's within an item stratum for the W population. Then using a technique developed by Nathan Keyfitz to maximize the overlap ELI's between populations and maintain the correct probabilities of selection, a second selection of ELI's for the all urban consumer unit population (U) was made proportional to the relative expenditures of the U population.

Each ELI was defined to be in one and only one POPS category so that the integration of the ELI sample and the outlet sample was by the ELI/POPS category concordance. Thus, the selection of ELI identified which (POPS) categories were to be used for outlet selection.

The following approach was used for outlet selection. Make a systematic selection of outlets reported for a given POPS category for the W population where the measure of size for each outlet was proportional to the average daily expenditure reported for each outlet by all households of the W population. The outlets for the U population were then selected by using the Keyfitzing technique.

For each ELI, the selection of a specific store item by a data collector is done using multi-stage probability selection techniques. A specific item is chosen in an outlet by sampling proportional to the percentage of sales of the kinds of items in the ELI. The four techniques used to obtain these percentages were, having the respondent furnish the percentages from records or memory, having the respondent rank the categories and then assign predetermined percentages to the rankings, using shelf space to estimate the percentage of sales or using equal probability if all else fails.

These procedures make possible an objective probability sampling of quotes throughout the CPFR. They also allow broad definitions of ELI's so that the same tight specification need not be priced everywhere. The wide variety of specific items greatly reduces the within EC item variance and allows a substantial reduction in the number of quotes required to obtain the same reliability as the current index. The sample is from 85 PSU's, with 2300 food stores and 22,400 retail outlets. Approximately 58,300 food quotes are priced each month and 56,000 commodity and services quotes are priced each month. Food, Gasoline, Housing and Utilities are priced monthly with most of the remaining items priced bimonthly. For the five largest pricing areas, all quotes are priced monthly, for all other self-representing publication areas, items not included above are priced bimonthly.

B. Rent (U.S.A.)

The sample of housing units prior to 1970 for the rent component of the CPI was based upon census data and other sources. The sample selection procedure was accomplished in several stages. First, a sample of housing "segments" for the 85 CPFR PSU's was selected from Census enumeration districts and other sources. A segment is a small geographic area containing up to 100 housing units, such as a city block, or a single apartment building, or even just certain floors in an apartment building. The segments were selected from a stratification of the block areas that made-up the housing universe. This stratification was based upon tenure (percentage of renter occupied), structure (percentage of single family units) and income.

The next step was to list each housing unit in a segment and then systematically sample from the listed units. The units selected were then screened to determine tenure. The eligible renter units were systematically sampled and selected units were initiated.

The 6,422 segments selected for the rent survey contained approximately 240,000 housing units. After subsampling 70,000 units were available for screening. Of the screened units, about 24,000 units were selected and about 21,000 were initiated.

For any given city, the sample of rental units is divided into six "panels" of approximately equal size, with one panel interviewed each month. Thus, one-sixth of the units are interviewed each month so that there is a steady flow of information on rent data from each city on a month-to-month basis.

The rent sample design provides for the collection of a current rent and the previous month's rent for the same interview. Since data is collected for both the current and previous month, it is possible to calculate price change over both the one-month and six-month time period. Furthermore, taking a special weighted average of these relatives defines a composite estimator. This composite estimator enables the measurement of rent to be an up-to-date measure of rent change, not a lagged estimation as when using only a six-month estimate of change, as in the current CPI. In addition, it is cost effective since fewer households need to be interviewed to obtain the same reliability.

The sample of screened units identified as owner-occupied or vacant for sale provided the sampling frame for the property tax survey. The sample was selected from eligible units such that on the average, about 2.5 owner housing units were selected per segment. This resulted in a sample of about 17,500 housing units.

The new construction universe for the rent and property tax sample is obtained from two sampling frames. The first frame represents the building permit issuing universe for the 85 PSU's. This frame was constructed and sampled by the Census Bureau for BLS. The second frame, for non-permit issuing areas, is constructed partly from the revised rent sample et partly by canvassing those in-scope non-permit issuing areas of the 85 PSU's. The same rent and property tax relatives are used for both population indexes in the CPFR.

The use of probability selection within an outlet was very successful in providing a proper distribution of types of items within an ELI at the national level. The use of this procedure made substantial gains in efficiency of the sample which was reflected in a substantial reduction of the sample size for the same reliability. The cost of the retail initiation and systems design was \$ 20 million dollars with a professional staff of 230 permanent employees and field personnel. The CPI is published monthly and further information is available on request.

Further information concerning the Consumer Price Index Revision may be obtained by writing to Miss Barbara A. Boyes, GAO Building, 441 G. St., N.W., Washington D.C. 20212 or by telephoning 202-523-1098.

STUDY ON INCOMPLETE DATA (U.S.A.)

A study on Incomplete Data (schedule or item non-response or unusable schedules or items) in sample surveys has been undertaken by the Committee on National Statistics of the National Academy of Sciences-National Research Council. A Panel with Professor Olkin of Stanford University as chair-person has been appointed to make the study.

The Panel expects to publish :

- a collection of case studies of interesting treatments of incomplete data
- a review of theory, including both the theory of current practice and ongoing research in the area
- a listing of computer programs concerned with incomplete data, including imputation as well as adjustment
- an annotated bibliography of papers, reports, and books concerned with incomplete data.

The Panel and staff would be very interested to learn of :

- a) possible case studies ; especially, but not limited to, case studies in which empirical validation of the methods used in dealing with incomplete data has been possible either through the availability of incomplete data or summary statistics whether from matching studies, other surveys or any other source
- b) ongoing research both theoretical and applied, so that it can at least mentioned in the reports that will appear in about 18 months
- c) any computer programs that exist or are being developed that deal with incomplete data
- d) bibliographies in process of publication or preparation
- e) information on trends over the past 25 years in the volume and composition of incomplete data as found in survey organizations and continuing surveys.

Please send information on these subjects to : William G. Madow, Committee on National Statistics, National Research Council, 2101 Constitution Avenue, NW, Washington, D.C. 20418.

From W.G. Madow
Statement for AMSTAT News.

SURVEYING THE EXTENT OF DRUG USE (Australia)

Early in 1977 the South Australian Government appointed the Royal Commission into the Non-medical Use of Drugs to enquire into the use of narcotic, analgesic, sedative and psychotropic drugs or substances of dependence, excluding alcohol or nicotine. Essentially the commission was to marshal from available sources, both local and overseas, information concerning such drugs, the extent and character of the use and abuse of drugs, and the medical, social and economic factors associated with such practices.

Many submissions were placed before the Commission and estimates of the extent of drug use, especially of illicit drugs, varied considerably. It became obvious that an independent estimate of the prevalence of drug use would be needed to compare these submissions. Various attempts have been made to measure drug use, especially of the so-called "soft" drugs (marijuana, analgesics, etc.), but most of these were of special populations and not directly of assistance to the Commissioners.

The Commission initiated discussions with the Adelaide office of the Australian Bureau of Statistics in August 1977 to determine if any assistance could be provided. Strong emphasis was placed on the need for results to be available for collation and analysis by about July 1978, and it was decided that in the absence of useful data, a special survey would be undertaken.

QUESTIONNAIRE DESIGN

The survey was intended to determine the extent of drug use, both for medical and non-medical purposes, awareness of drugs and opinions about possible changes to the laws regarding illicit drug use. The most recent large scale enquiry into drug use was carried out by the National Commission on Marijuana and Drug Abuse in 1972-3. Information from the U.S. Commission together with data from smaller scale Australian surveys was used to assist in the design of this survey.

Several draft questionnaires were tested in small discussion groups and were administered to volunteer groups by trained ABS interviewers and by Commission staff. After considerable revision, the questionnaire was tested in the field to determine the interpretation of questions by more typical respondents, and to disclose weaknesses in expression. In addition, it was anticipated that the tabulations of results from pilot tests would provide a guide to the levels of drug use to be expected in the main survey. Subsequently, two pilot surveys were undertaken, the first of 403 dwellings, the second of 614 dwellings; these were selected at random.

Because many of the questions involved disclosure of information of a highly personal nature, it was necessary to obtain an interview with each respondent individually. The usual residents of each household were recorded in two age groups, 13-17 years of age and 18-60 years of age and a different questionnaire form was to be used for each group. The Youth Questionnaire, coloured green for easy identification, was for the younger group and omitted questions regarding changes to the law, but was otherwise identical to the adult questionnaire (coloured white).

Questions were grouped into eight sections, the first being a "lead in" to the questionnaire proper, and requesting answers to general questions only. Following sections in order were :

1. Proprietary "over the counter" drugs
2. Sedatives and Painkillers
3. Marijuana
4. Tranquilisers and anti-depressants
5. Amphetamines
6. Questions relating to the existing laws and penalties regarding illicit drug use, and opinions on the effects of use of drugs
7. Demographic information.

The questions in sections 2, 3, 5 and 6 were very similar and designed to elicit information about medical and non-prescription use of the respective drugs. To aid recognition, the respondent was handed a card containing colour photographs of the main drugs in each group before being questioned. The interviewer read a short preamble explaining the uses of the drug, and then asked about the type of use (on prescription or not), when the drug was used first, the most recent use, and current use. Respondents were requested to nominate if they considered themselves to be frequent or just occasional users.

Each of the sections contained 12 questions, five of which were mandatory, the rest dependent on response to earlier questions. These groups of questions were (unavoidably) repetitive, and the order in which the questions were asked was varied between groups; this together with the prompt cards was intended to maintain the respondent's attention.

Because possession of marijuana is a criminal offence in South Australia, disclosure of use of the drug may be inhibited. Various strategies were considered to collect this information, including a randomised response technique which was tested in one pilot survey. It was found that this technique, which was designed to reduce evasive answer bias, did not significantly increase the number of affirmative responses to the controversial question, and was rather time-consuming to administer.

It was finally decided to use a separate self-administered questionnaire containing 12 questions. The interviewer was to hand the form together with an envelope and a pen to the respondent, explain that each question required an answer (regardless of the respondent's knowledge of the drug), and seal the form in the envelope.

Interviewers did not see the answers, as the forms were related to the rest of the questionnaire by Commission staff later.

Questions of opinion of the existing laws regarding possession and use of illicit drugs proved to be difficult to design. Prompt cards were used to suggest alternative penalties to reduce the number of categories into which an answer could fall. The respondent was not told the present penalties (except that it was an offence to possess drugs) unless they requested further information. It was intended that this would reduce any bias towards selection of the current penalties. Only two drugs were considered in this part of the questionnaire, marijuana and heroin, the latter indicative of attitudes to lesser known "hard" drugs.

The final set of questions was to determine occupation, academic qualifications and income. These attributes were classified as in the 1976 census to enable comparison. However, it is intended to reclassify occupations by skill criteria.

The questionnaire form was designed to be self-coding, with the exception of the occupation and qualifications questions.

Sample design

The selection of individuals for interview was a two stage process. A clustered stratified random sample of dwellings was selected by ABS as the first stage, and the interviewer was to select randomly, from the usual residents of each dwelling, persons for interview.

A sample of 1.25% of dwellings in the Adelaide Statistical Division was selected, the size being determined by the estimated standard error of the cells of lowest frequency. The cell sizes were estimated from pilot survey counts, and from other studies of drug use. The sample comprised approximately 4,100 dwellings, and, given the secondary sampling scheme below, plus sample loss due to refusals, non-contacts, etc., yielded about 3,000 interviews.

Although the cost per interview was increased, it was decided to use a 1.25% sample of dwellings to obtain a wide cross-section of respondents, but to limit the number of interviews carried out in each dwelling. In the pilot test, the average time to complete an interview was 40 minutes, and a limit of two interviews per dwelling would reduce the demand on respondents (and interviewers) to a reasonable level.

Selected dwellings were then randomly assigned a Type A or Type B classification. Interviewers were instructed to list all the usual residents of the dwelling in age order, from youngest to oldest. As they were listed each person who was eligible for interview was automatically numbered 1-9 within age strata. In Type A dwellings, persons 13-17 and 18-34 years of age were eligible for selection, and received a number; in Type B persons 13-17 and 18-60 years were eligible. As the numbers of A and B Types were equal, persons 35-60 had a probability of selection in only half of the dwellings. Each interviewer was issued with a table of numbers 1-9 randomly listed, and these were used to select each individual to be interviewed. Each person in each eligible age stratum had an equal probability of selection.

The total number of dwellings surveyed was 4,030; they were divided into 116 workloads. Most interviewers employed by the Commission were experienced in administering questionnaires of a similar format, and were familiar with the rules regarding updating the sample selections. In order to reduce non-sampling bias, training courses covering the unique aspects of this survey, interviewing techniques, etc. were conducted. In addition, interviewers were field-checked to ensure that information was collected in a consistent manner.

After excluding households found to be outside the scope of the survey, sample loss was 10.8% and refusals were 3.2% of dwellings contacted. These results were considered to be satisfactory as the survey was voluntary and covered rather controversial characteristics. Refusal rates in the pilot surveys were 5.0% and 5.8% respectively; these higher rates are possibly due to the extensive publicity given to the main survey.

Another factor contributing to the high response rate is believed to be the use of primary approach letters. A letter was sent to each dwelling advising that the dwelling had been selected in the survey, and explaining the aims of the Commission. While it is possible that the use of letters enabled the respondent to avoid the interviewers (who called at least five times on each dwelling), this is not reflected in a higher non-contact rate. Interviewers were strongly of the opinion that approach letters aided them in obtaining their interviews, and overall the decision to deliver the letters in the week before the household was surveyed appears to have contributed to the high response rate.

Because persons in each age/sex stratum had a probability of selection which varied between households weighted records were produced. The weight for each record is calculated as follows. Let the total number of persons in the a^{th} age-sex cell in all selected households be $n(a)$, the total number of persons in the a^{th} age-sex cell in the total population be $N(a)$ and the number of persons in the i^{th} household in the a^{th} age-sex cell be $n(a,i)$. Then $WT(1) = N(a)/n(a)$ and $WT(2) = n(a,i)$. The appropriate weight for the i^{th} record will then be $WT = WT(1) \times WT(2)$.

The weight will vary between households and is dependent upon the number of persons eligible for selection in that household. It will automatically allow for the fact that respondents had different probabilities of selection in Type A or B dwellings. The weighted tape was used to produce tabulations of population estimates required by the Royal Commission.

Standard errors were calculated using a split-half method devised by ABS.

Results

As yet, only preliminary results have been released by the Commission. Estimation of the use of marijuana was one of the most controversial aspects of this survey, and a table of some of the estimates, together with relevant standard errors is given below. The table indicates that 21% of males aged 13 to 60 and 11% of females aged 13 to 60 have ever used marijuana.

TABLE : FIRST USE OF MARIJUANA v SEX

Estimated numbers aged 13 to 60 in the Adelaide Statistical Division (numbers in brackets are % standard errors)

| First use of Marijuana | Males | Females | Persons |
|---------------------------|----------------|----------------|----------------|
| Used at first opportunity | 42 700 (6.8) | 21 400 (9.7) | 64 100 (5.6) |
| Used later | 19 800 (10) | 10 500 (14) | 30 300 (8.1) |
| Total ever used | 62 500 (5.7) | 31 900 (7.9) | 94 400 (4.5) |
| Never used | 224 500 (2.9) | 250 700 (2.8) | 475 200 (2.0) |
| No answer | 12 600 (13) | 15 500 (12) | 28 200 (8.4) |
| TOTAL | 299 600 | 298 100 | 597 700 |

from Tim GOODE
(Australian Bureau of Statistics)
and Warwick HEINE
(S.A. Royal Commission into
Non-medical Use of Drugs)

Paper issued in the Newsletter,
August 1978, n° 5, Statistical Society
of Australia

TRAINING

THEORETICAL PREPARATION AND METHODS OF SELECTION OF SURVEY STATISTICIANS : SPANISH EXPERIENCE

"To study the phenomenon that different categories of statisticians are often unaware of, and desinterested in, each other's professional achievements, and apparently lack adequate appreciation for the interrelationship of their respective work".

1. Introduction

The above sentence corresponds to the first point of the terms of reference for the ISI Committee on Integration of Statistics (1976). I think that, at least in Spain, if the word "statistician" were substituted by "survey statistician", the sentence would still hold.

In this note, we shall try and analyze the problem for two categories of survey statisticians :

- a) Those responsible for the sample in its phases of frame, hierarchy of units, charts and lists, stratification, size of sample, estimation, sampling and non-sampling errors, etc.
- b) Those who direct field work, select and train interviewers, take an active part in the design of questionnaires, control and inspect the collection of data, direct the coding and editing, etc.

We shall take the first task "sample design" and the second "field work" both being integrated in the "survey design".

2. Spanish experience

In Spain both categories of survey statisticians are taken from Government Statisticians with higher university degree and are selected by the Instituto Nacional de Estadística by competitive examinations among Economists and Mathematicians. The knowledge of sampling theory required is the same for both cases. The problem therefore does not arise at this stage.

Difficulties begin after a few years of carrying out the specific work of each group. In my note to the Chairman of the Committee previously mentioned I said : "Whilst, for example, in the academic class, other statisticians are often considered as 'practitioners', these, for their part consider as 'theoreticians' not only the former but also their own colleagues who try to make official statistics more scientific". I think that this sentence referred to statisticians in general still holds for the two categories of statisticians considered here.

The separation of these two categories of survey statisticians may become quite dangerous, if those considered as "theoretical minds separated from reality" and those responsible for field work come under different departments, as happened in our office up to 1972 when both units were integrated into the present Subdirección General of Sampling and Censuses.

2.1 Minimum knowledge necessary for survey statisticians to carry out their work

a) For sample design

In accordance with the phases mentioned in the introduction, the minimum knowledge shall comprehend the following concepts : the frame, stratification, determination of size of sample, selection and estimation processes, sampling and non-sampling errors

b) For the field work

Programming ; estimation of requirements of personnel ; workload for interviewers ; time table ; elaboration and control of budgets ; preparation of tests for selections ; preparation of handbooks ; permanent training in continuous surveys ; etc.

2.2 Sampling theory required of all statisticians to enter the Instituto Nacional de Estadística

During the last twenty years the Instituto Nacional de Estadística has announced an average of ten vacancies per year of statisticians with master university degree. Most of the candidates were economists or mathematicians, and there are approximately ten candidates for each post.

The selection exam consists in various exercises with a total of 167 themes. Out of these, 18 belong to sampling theory. The candidate must develop one, chosen at random, on the blackboard within a maximum of thirty minutes.

Out of the candidates that pass, about 10 % are absorbed by the sampling and censuses Subdirección General.

2.3. Training and updating of survey statisticians

As we have already mentioned the initial knowledge about sampling theory is the same for the survey statisticians dealing with sample design or with field work. It is practice and further studies which widen the gap.

Perhaps some seminars as well as refresher courses and updating carried out with adequate periodicity and compulsory attendance of all survey statisticians would be an appropriate means to fill up this gap.

In our Office, even if some seminars and lectures have taken place in the past, these were neither systematic nor compulsory.

2.4. Programs of sampling theory at the University

I shall mention now, as an example, the programme of sampling theory explained by myself to the students of the fourth course in the Faculty of Economic Sciences of the Universidad Autónoma de Madrid during 9 months and for three hours per week :

1. Population, frame and sample. Probability sampling. Sampling distribution. Selection with and without replacement. Probability that the u_i unit is in the sample. Precision and accuracy of an estimator. Bias of selection and estimation and its effects.

2. Linear unbiased estimators : application to the mean, total and proportion. Variances of these estimators in sampling with and without replacement. Expected value of the sample variance. Estimation of the variances.

3. Stratified sampling. Estimators for the mean, total and proportion. Expected value and variance, and optimal. Gain in precision.

4. Systematic sampling. Estimators and variances. Approximate sampling errors. Variance of the estimator with independent samples. Estimation of the variance.

5. Estimation of the change between two occasions. Estimator of minimum variance. Rotation of the sample with partial overlapping. Mean over two occasions.

6. Ratio estimators. Bias of the ratio estimator. Approximate variance. Simple and combine ratio estimators in stratified sampling.

7. Regression estimators. Bias of the regression estimator. Variance with preassigned value of b. Minimum variance. Comparison of variances. Regression estimators in stratified sampling.
8. Double sampling. Application to stratified sampling. Optimum allocation in double sampling.
9. Application of double sampling to ratio and regression estimators.
10. Sampling of compact clusters. Intra cluster correlation. Optimum cluster size.
11. Sample size determination. Design effect. Rare items. Relationship between n and N for a fixed variance.
12. Cluster sampling with subsampling. Estimators and variances. Optimization.
13. Sampling with replacement and probabilities proportional to size. Lahiri's method. Estimator of Hansen and Hurwitz. Extension to multistage sampling.
14. Sampling without replacement and unequal probabilities. Estimator of Horwitz and Thompson. Extension to multistage sampling. Particular case of equal probabilities.
15. Variance of a linear estimator : Theorem I of Durbin. Particular cases : a) with replacement and first stage units with probabilities proportional to size ; b) without replacement and first stage units with probabilities proportional to size.
16. Variance of a linear estimator : Theorem of Madow. Applications.
17. General method for the estimation of variances : Theorem II of Durbin and practical rule. Ultimate cluster approach.
18. Non-response : causes and effects. Treatment of non-responses : adjustment of data ; method of Hansen and Hurwitz ; method of Politz and Simmons.
19. Random response model.
20. Quality of the statistical data. Model of Hansen, Hurwitz and Bershad. Indicators of quality.

2.5. Resources and production of Statistics

The Instituto Nacional de Estadística is in charge of the large surveys : Active Population, with 60,000 households quarterly, Family Budget, with 25,000 households every five years, Consumption with 2,000 households quarterly, Equipment and cultural level, Fertility, etc.

On the other hand, the Institute includes within the structure of the multipurpose General Population Design, other surveys, with practically no extra cost, closely related to the Active Population Survey, such as Educational Expenses, Reading Habits, Sports, etc.

The budget of the Institute allocated for censuses and sample surveys is roughly seven millions US dollars, approximately 25 % of the total budget of the Institute.

Among the population surveys, the higher cost corresponds to the Active Population Survey which is about two and a half million dollars per year. Nearly five hundred people participate in this survey out of the two thousand staff of the Institute.

From what has been said above, it is clear that the active role of sampling in the statistical activities of our Central Office and the substantial part of the budget allocated to these tasks require a sound theoretical basis and an adequate training.

From J.L. SANCHEZ-CRESPO
Instituto Nacional de Estadística
Universidad Autónoma de Madrid

TERMINOLOGY

IASS Committee on Terminology. Three members (F. AZORIN, J.L. SANCHEZ-CRESPO, and L. KISH) met in Mexico on November 6-10, 1978. We agreed that publication in the Survey Statistician of proposals for terminology would be most beneficial at this stage. Also that the following would be a good start : "Populations for Survey Sampling".

POPULATIONS FOR SURVEY SAMPLING

We need four distinct populations for the practice of survey sampling :

- survey population, actually observed in the sample
- frame population, from which the sample was initially selected
- target population
- inferential populations.

I believe that these four are all needed, but also that these four will serve our needs. Commonly only two populations have been contrasted ; and the term "target population" has been used at times for frame population in contrast to the survey population of the actual sample, but at other times it seems to mean the population(s) to which inference is made in contrast to the frame population. I am trying for a good compromise with both the literature and our needs in practice. The use of "frame population" has not suffered much inconsistency.

The differences between diverse populations may be defined with terms fairly well accepted in survey literature ; also illustrated with two kinds of surveys : first a population of household members, and secondly persons with a specified disorder from a sample of hospitals.

Survey populations differ from frame populations in two important respects. First, there are losses due to non-responses ; eligible members who are not observed, due to refusals, not-at-homes, not found, etc. Non-response rates vary greatly but some loss is unavoidable. On the contrary, differences due to biased selection from frames are avoided by skilled samplers, but are frequent sources of biases in many survey samples. Overcoming potential biases in sampling frames is the sampler's first task. The sample cases should be compensated (when not self-weighting) with inverse weights for inequalities of selection probabilities ; and adjustments for non-responses may also be introduced. These should bring the sample closer to the frame population. Nonresponses occur mostly at the household level in those surveys ; but in a survey of hospitalized patients, nonresponses can occur from both hospitals and patients. In general, each stage of sampling units is a potential source of nonresponse.

The differences between frame and target populations are mostly due to coverage errors. These can be overcoverage (when the frame includes cases which do not belong to the target population and are difficult to weed out) but most of our troubles are due to undercoverage, to incomplete frames. Listings of dwellings in blocks, and of persons in households are notoriously incomplete. Listings of hospitals, and of patients will seldom be 100 per cent complete. Also lists tend to be more or less incomplete by the time the survey is taken. These kinds of incompleteness are usually difficult to assess and to compensate.

Deliberate exclusions of well defined subpopulations belong to the distinction between target and inferential populations, I believe. The target population is defined by the aims of the sampler and of the sample design. The inferential populations change with the aims of the analysts and of the users of the data. Samples of the USA often exclude Alaska and Hawaii for reasons of unit costs. Household surveys exclude institutional and military populations. Hospital surveys may exclude disorders treated outside hospitals, plus the untreated and dead, for reasons of costs. In addition to such exclusions, the analyst may also want to include other populations to broaden his inference. Most likely he also extrapolates to years beyond the target period. This may be easy with a stable population, but many populations, such as persons with some disorder, may be far from stable. The process of inference must often depend on scientific knowledge and personal judgment. Clearly there are many possible inferential populations.

We used the singular for the survey, the frame and the target population. But every survey is multipurpose and is used for many subpopulations included in the survey population. Some of these subclasses may be planned for as domains of study, e.g. regional domains ; other subclasses like age and sex of members just cut across the entire design (crossclasses). In addition to population defined by extent, the same survey may have different contents as well, e.g. persons, families and households from the same sample.

The four populations named above ascend from means to ends, as one moves from survey to frame to target to inferential populations. The link from frame to target should be made close for obvious reasons. The limitations of survey resources influence the choice of frames. The choice of the target population should conform to guesses about inferential populations in the future.

from Leslie KISH
University of Michigan

PAPERS

THE SWEDISH PARTY PREFERENCE SURVEYS

During 1968-1972 the National Central Bureau of Statistics (SCB) received grants for research within the area of political opinion polls and for conducting a number of pilot studies. In November 1972 the SCB launched their regular Party Preference Survey (PSU) which since has been carried out every May and November. Election years a survey has also been conducted in February. The object of the surveys is to estimate the results of a general election at the time of the survey, and the strength of the political parties in different subclasses of the electorate. The reasons for giving grants to the SCB for conducting such surveys are that they not only stimulate and give a base for the political debate both in mass-media and among the general public, but also that they constitute a base for decision-making among political parties, mass-media and the general public. The PSU can also give valuable information for political science research. Moreover, the PSU provides an alternative to the polls regularly published by private institutes.

In Sweden the civil service is not connected with the political party or parties in power and is looked upon as unbiased and neutral of the political parties. By giving grants to the SCB, a political opinion poll was created, free of any commercial interests and offering full public insight in the production of the estimates and the underlying research work. The results are available to all interested, regardless of economic means.

Sweden has the advantage of having a central register of the whole population which can be used as a sampling frame. The register contains information that permits persons not belonging to the electorate to be excluded, viz. persons that are not Swedish citizens, persons under legal disability and persons under age.

In spite of being drawn systematically (by computer), the sample is considered as a simple random sample composed of three rotating panels, each holding approximately 3 000 persons. Consequently about 9 000 persons are to be interviewed in a PSU - 3 000 for the first, 3 000 for the second and 3 000 for the third and last time. The sampling fraction of a panel is 1/1980. The panel approach and the great sample size provide opportunity to estimate changes and totals even in subclasses without getting too great sampling errors.

No extensive questionnaire is required for the PSU interview. The socio-economic classification excepted, all the background variables are obtained from the sampling frame or from other registers. There remain only a few relevant research variables which can be classified into three categories :

- questions concerning what political party the respondent has the greatest preference for
- questions concerning the voting at the latest general election
- questions concerning the respondent's voting intentions in a nearby hypothetical election.

In order to get more information, respondents who have answered "don't know" or have refused to give an answer about party preference are asked what "party block" - non socialist or socialist - they prefer or had voted for.

The socio-economic classification requires several questions concerning current or previous occupation.

The interview data are mainly collected during a fortnight through telephone contacts by local SCB interviewers. Some follow-up interviews are made by a group of centrally stationed interviewers. About 95 % of the interviews are carried out by telephone. If the respondent can not be reached by telephone or refuses a telephone interview, the interview can be carried out in a visit at home.

In two pilot studies a secret ballot approach in the PSU has been tested. However, these studies were not conclusive and gave no unambiguous differences in the tested distributions as compared with telephone interviews (see : Staffan Sollander, "Experiments with anonymous answers to PSU-questions" (1975, in Swedish) and "Further experiments with anonymous answers to PSU-questions" (1976, in Swedish), SCB, Stockholm). Besides, the secret ballot approach requires at-home interviews which are approximately three times as expensive as telephone interviews. Moreover, if the secret ballot approach is conducted in a proper way it is impossible to give estimates of the gross changes between successive surveys and of distributions in subclasses defined by register-data as it then is impossible to identify a respondent's answers.

The non-response due to refusal is normally about 7 % of the sample, due to not-at-home about 3 % and due to illness, etc. about 2.5 %. In addition, there are always some respondents who refuse to answer some of the questions.

The estimates of the results of a hypothetical election at the time of the interviewing are made by post-stratification. The variables used for post-stratification are behaviour at the latest election (six party strata, non-voters and too young to vote at the latest election) and regions of constituencies (ten strata). In all there are 8 x 10 post-strata.

To each sample unit a "probability of participation" is assigned based on the answer to a question about intentions to participate in a hypothetical election. A similar question has been asked in surveys immediately before real elections and as the electoral registers are public it has been possible to compare the intentions with the actual behaviour and hence to calculate the "probability of participation". For example, 97 % of those who in the survey before the 1976 general election answered that they would participate, really did, as well as 41 % of those who said that they had no intention to participate.

As the "probability of participation" varies between the subgroups also the non-response has to be included in the estimator as the non-respondents normally have a lower "probability of participation" than the respondents. Thus, in the last summation of the denominator and nominator of the estimator respectively, the non-response is distributed within each region stratum according to the distribution of the respondents in the region stratum.

The estimator of a political party's proportion of the electorate in a hypothetical general election is

$$p_i = \sum_k \sum_j W_{kij} \frac{\sum_j x_{kij} y_{kij} q_{kij} + \sum_j x_{kij} (1 - y_{kij}) \frac{\sum_j x_{kij}}{\sum_j y_{kij}}}{\sum_j y_{kij} + \sum_j (1 - y_{kij}) \frac{\sum_j y_{kij}}{\sum_j y_{kij}}}$$

where

- $x_{kij} = \begin{cases} 1 & \text{if the } j\text{'th sampling unit belongs to the } i\text{'th "election acting" stratum and the } \\ & \text{h'th region-stratum and has the intentions to vote for the party} \\ 0 & \text{otherwise} \end{cases}$
- $y_{kij} = \begin{cases} 1 & \text{if the } j\text{'th sampling unit belongs to the } h\text{'th region stratum and has stated for which} \\ & \text{party he intends to vote and that h belongs to the } i\text{'th "election-acting" stratum} \\ 0 & \text{otherwise} \end{cases}$
- $q_{kij} =$ probability of participation for the j 'th sampling unit belonging to the h 'th post-stratum
- $W_{hi} =$ the weight of the h 'th "election-acting"-region stratum . $\sum_k \sum_j W_{kij} = 1$

The variance of this estimator is not calculated. Instead the variance of a somewhat simplified estimator is calculated.

Special estimates are given of the "party-blocks". These are made by a post-stratification estimator similar to the one above.

The distribution of party preferences as well as the variances are estimated with the usual estimator for simple random sampling.

About six weeks after the start of the interviewing the results of the survey are presented in an extensive press release. A month later, the official report is published in the SCB series of Statistical Reports.

In the press release the results of a hypothetical election are published as estimated 95 % confidence intervals. We also publish estimated confidence intervals of the changes that have taken place from the previous general election and from the previous PSU. The idea behind the presentation of confidence intervals is to discourage inferences from too small changes and differences. Probably this has contributed to more prudent comments of poll results in the mass-media.

Tables of the flows between the political parties since the previous election and since the previous PSU are also published. This gives us the opportunity to analyse both net and gross flows of the voting preferences.

The great sample size makes it possible to estimate party preference distributions in several subclasses. The subclasses are determined by age, sex and age, marital status and age, number of children and age, respondent's income, total income of respondent and spouse, relative number of population in localities in the community of registration, constituencies and socio-economic conditions (SEI-classification). The basis for the SEI-classification are answers to questions about occupation. Employees are classified into seven classes defined by the education normally required for the occupation in question. Entrepreneurs and farmers are classified into three classes, determined by the number of employees and acreage respectively. Retired persons are classified by earlier occupation, and married but not gainfully occupied persons by occupation of the spouse.

At the start the party preference distributions for subclasses were published as estimated 95 % confidence intervals in such a way that - including four classes of non-response - the sum of the midpoints were one hundred per cent. The intentions were to point out the effects of sampling errors and non-response. However, when the non-response was unequal in the subclasses the presentation manner turned out to complicate the analyses of differences and time series. Consequently these distributions are now published as point estimates plus/minus half the 95 % confidence interval, with the sum of the party estimates amounting to one hundred per cent. The estimates from the preceding PSUs are given in the same table as for the current PSU. In the same table as the party estimates the figures of non-response are given. The tables are printed by the computer in a way that makes it possible to use them as originals for printing. An example of the tables are given in table 1.

Table 1. The electorate distributed by party preference as well as by age at the end of the interview year and sex.

The estimates are percentages \pm twice the standard deviation. The interval that is formed constitutes a 95 per cent confidence interval, i.e., an interval which with 95 % probability contains the correct population value if there are no systematic errors.

| | With party preference % | | | | | | The whole net sample % | | | | | Base figures | |
|--------------------------------|-------------------------|----------------|-----------------|------------------|---------------|---------------|------------------------|--------------------------|----------------|----------------|---------------|--------------|------|
| | Centre party | Liberal party | Conservative p. | Social Democrats | Communists | Others | With party preference | No preference Don't know | Refused | Not at home | Total | | |
| Men aged 18-34 | | | | | | | | | | | | | |
| Nov 76 | 22,6 \pm 2,0 | 9,8 \pm 1,4 | 12,5 \pm 1,6 | 47,1 \pm 2,4 | 5,9 \pm 1,1 | 2,1 \pm 0,7 | 100. | 72,8 \pm 2,3 | 11,3 \pm 1,6 | 9,6 \pm 1,5 | 6,4 \pm 1,2 | 100. | 1497 |
| May 77 | 19,0 \pm 1,9 | 7,8 \pm 1,3 | 12,6 \pm 1,6 | 52,4 \pm 2,5 | 6,5 \pm 1,2 | 1,7 \pm 0,6 | 100. | 72,6 \pm 2,3 | 12,1 \pm 1,7 | 8,0 \pm 1,4 | 7,3 \pm 1,4 | 100. | 1424 |
| Nov 77 | 19,5 \pm 1,9 | 8,1 \pm 1,3 | 13,2 \pm 1,7 | 50,8 \pm 2,5 | 6,9 \pm 1,3 | 1,5 \pm 0,6 | 100. | 75,0 \pm 2,3 | 12,0 \pm 1,7 | 7,3 \pm 1,4 | 5,7 \pm 1,2 | 100. | 1404 |
| May 78 | 16,7 \pm 1,8 | 7,3 \pm 1,3 | 14,1 \pm 1,7 | 53,2 \pm 2,5 | 6,9 \pm 1,3 | 1,8 \pm 0,7 | 100. | 74,6 \pm 2,3 | 12,9 \pm 1,8 | 7,2 \pm 1,4 | 5,4 \pm 1,2 | 100. | 1398 |
| Men aged 35 and above | | | | | | | | | | | | | |
| Nov 76 | 21,4 \pm 1,4 | 10,4 \pm 1,1 | 14,1 \pm 1,2 | 51,0 \pm 1,7 | 2,0 \pm 0,5 | 1,1 \pm 0,4 | 100. | 75,9 \pm 1,6 | 7,7 \pm 1,0 | 11,8 \pm 1,2 | 4,7 \pm 0,8 | 100. | 2859 |
| May 77 | 20,3 \pm 1,4 | 7,8 \pm 0,9 | 16,5 \pm 1,3 | 52,8 \pm 1,7 | 1,9 \pm 0,5 | 1,2 \pm 0,4 | 100. | 74,9 \pm 1,6 | 8,3 \pm 1,0 | 10,2 \pm 1,1 | 6,6 \pm 0,9 | 100. | 2940 |
| Nov 77 | 21,3 \pm 1,4 | 8,2 \pm 0,9 | 15,6 \pm 1,2 | 51,5 \pm 1,7 | 2,1 \pm 0,5 | 1,3 \pm 0,4 | 100. | 76,5 \pm 1,5 | 7,7 \pm 1,0 | 9,7 \pm 1,1 | 6,1 \pm 0,9 | 100. | 2899 |
| May 78 | 19,3 \pm 1,3 | 9,1 \pm 1,0 | 15,4 \pm 1,2 | 52,8 \pm 1,7 | 2,2 \pm 0,5 | 1,2 \pm 0,4 | 100. | 76,4 \pm 1,5 | 8,9 \pm 1,0 | 8,9 \pm 1,0 | 5,7 \pm 0,8 | 100. | 2984 |
| Women aged 18-34 | | | | | | | | | | | | | |
| Nov 76 | 25,1 \pm 2,3 | 12,9 \pm 1,8 | 9,6 \pm 1,5 | 42,9 \pm 2,6 | 6,7 \pm 1,3 | 2,8 \pm 0,9 | 100. | 70,7 \pm 2,4 | 11,3 \pm 1,7 | 13,6 \pm 1,8 | 4,4 \pm 1,1 | 100. | 1325 |
| May 77 | 22,6 \pm 2,2 | 9,4 \pm 1,5 | 7,4 \pm 1,4 | 50,9 \pm 2,6 | 6,8 \pm 1,3 | 2,9 \pm 0,9 | 100. | 69,6 \pm 2,4 | 13,4 \pm 1,8 | 11,0 \pm 1,7 | 6,0 \pm 1,3 | 100. | 1354 |
| Nov 77 | 22,6 \pm 2,1 | 9,4 \pm 1,5 | 7,1 \pm 1,3 | 51,8 \pm 2,6 | 6,6 \pm 1,3 | 2,4 \pm 0,8 | 100. | 71,3 \pm 2,4 | 13,9 \pm 1,8 | 10,4 \pm 1,6 | 4,4 \pm 1,1 | 100. | 1353 |
| May 78 | 19,2 \pm 2,0 | 8,7 \pm 1,5 | 8,4 \pm 1,4 | 54,7 \pm 2,6 | 6,8 \pm 1,3 | 2,1 \pm 0,7 | 100. | 71,5 \pm 2,4 | 15,8 \pm 2,0 | 8,7 \pm 1,5 | 4,0 \pm 1,0 | 100. | 1332 |
| Women aged 35 and above | | | | | | | | | | | | | |
| Nov 76 | 23,9 \pm 1,5 | 11,4 \pm 1,1 | 14,4 \pm 1,2 | 47,0 \pm 1,7 | 1,1 \pm 0,4 | 2,3 \pm 0,5 | 100. | 66,8 \pm 1,6 | 10,0 \pm 1,0 | 18,1 \pm 1,3 | 5,2 \pm 0,8 | 100. | 3162 |
| May 77 | 21,9 \pm 1,4 | 9,2 \pm 1,0 | 16,3 \pm 1,3 | 49,9 \pm 1,7 | 0,7 \pm 0,3 | 1,9 \pm 0,5 | 100. | 67,3 \pm 1,6 | 11,4 \pm 1,1 | 15,1 \pm 1,2 | 6,2 \pm 0,8 | 100. | 3217 |
| Nov 77 | 22,9 \pm 1,4 | 10,3 \pm 1,0 | 15,0 \pm 1,2 | 49,0 \pm 1,7 | 0,9 \pm 0,3 | 1,9 \pm 0,5 | 100. | 68,6 \pm 1,6 | 10,8 \pm 1,1 | 14,1 \pm 1,2 | 6,5 \pm 0,9 | 100. | 3237 |
| May 78 | 21,1 \pm 1,4 | 11,5 \pm 1,1 | 15,3 \pm 1,2 | 49,4 \pm 1,7 | 1,1 \pm 0,3 | 1,8 \pm 0,4 | 100. | 69,3 \pm 1,6 | 11,1 \pm 1,1 | 15,6 \pm 1,2 | 6,0 \pm 0,8 | 100. | 3289 |
| Total men | | | | | | | | | | | | | |
| Nov 76 | 21,8 \pm 1,2 | 10,2 \pm 0,9 | 13,6 \pm 1,0 | 49,7 \pm 1,4 | 3,3 \pm 0,5 | 1,4 \pm 0,3 | 100. | 74,8 \pm 1,3 | 8,9 \pm 0,9 | 11,0 \pm 0,9 | 5,3 \pm 0,7 | 100. | 4356 |
| May 77 | 19,9 \pm 1,1 | 7,8 \pm 0,8 | 14,9 \pm 1,0 | 52,7 \pm 1,4 | 3,3 \pm 0,5 | 1,4 \pm 0,3 | 100. | 74,2 \pm 1,3 | 9,5 \pm 0,9 | 9,5 \pm 0,9 | 6,8 \pm 0,8 | 100. | 4364 |
| Nov 77 | 20,7 \pm 1,1 | 8,2 \pm 0,8 | 14,8 \pm 1,0 | 51,3 \pm 1,4 | 3,6 \pm 0,5 | 1,4 \pm 0,3 | 100. | 76,0 \pm 1,3 | 9,1 \pm 0,9 | 8,9 \pm 0,9 | 6,0 \pm 0,7 | 100. | 4303 |
| May 78 | 18,5 \pm 1,1 | 8,5 \pm 0,8 | 15,0 \pm 1,0 | 52,9 \pm 1,4 | 3,7 \pm 0,5 | 1,4 \pm 0,3 | 100. | 75,8 \pm 1,3 | 10,2 \pm 0,9 | 8,4 \pm 0,8 | 5,6 \pm 0,7 | 100. | 4382 |
| Total women | | | | | | | | | | | | | |
| Nov 76 | 24,3 \pm 1,2 | 11,8 \pm 0,9 | 12,9 \pm 1,0 | 45,7 \pm 1,4 | 2,9 \pm 0,5 | 2,4 \pm 0,4 | 100. | 67,9 \pm 1,4 | 10,4 \pm 0,9 | 16,8 \pm 1,1 | 4,9 \pm 0,6 | 100. | 4487 |
| May 77 | 22,1 \pm 1,2 | 9,3 \pm 0,8 | 13,6 \pm 1,0 | 50,3 \pm 1,4 | 2,6 \pm 0,4 | 2,2 \pm 0,4 | 100. | 68,0 \pm 1,4 | 12,0 \pm 0,9 | 13,9 \pm 1,0 | 6,1 \pm 0,7 | 100. | 4571 |
| Nov 77 | 22,8 \pm 1,2 | 10,0 \pm 0,9 | 12,6 \pm 0,9 | 49,8 \pm 1,4 | 2,6 \pm 0,4 | 2,1 \pm 0,4 | 100. | 69,4 \pm 1,3 | 11,7 \pm 0,9 | 13,0 \pm 1,0 | 5,9 \pm 0,7 | 100. | 4590 |
| May 78 | 20,5 \pm 1,1 | 10,7 \pm 0,9 | 13,2 \pm 1,0 | 51,0 \pm 1,4 | 2,8 \pm 0,5 | 1,9 \pm 0,4 | 100. | 69,3 \pm 1,3 | 12,5 \pm 1,0 | 12,8 \pm 0,9 | 5,4 \pm 0,7 | 100. | 4621 |
| The whole electorate | | | | | | | | | | | | | |
| Nov 76 | 23,0 \pm 0,9 | 11,0 \pm 0,6 | 13,2 \pm 0,7 | 47,8 \pm 1,0 | 3,1 \pm 0,4 | 1,9 \pm 0,3 | 100. | 71,3 \pm 0,9 | 9,7 \pm 0,6 | 13,9 \pm 0,7 | 5,1 \pm 0,5 | 100. | 8843 |
| May 77 | 21,0 \pm 0,8 | 8,5 \pm 0,6 | 14,3 \pm 0,7 | 51,5 \pm 1,0 | 3,0 \pm 0,3 | 1,8 \pm 0,3 | 100. | 71,0 \pm 0,9 | 10,8 \pm 0,6 | 11,6 \pm 0,7 | 6,4 \pm 0,5 | 100. | 8935 |
| Nov 77 | 21,8 \pm 0,8 | 9,1 \pm 0,6 | 13,7 \pm 0,7 | 50,6 \pm 1,0 | 3,1 \pm 0,4 | 1,7 \pm 0,3 | 100. | 72,6 \pm 0,9 | 10,5 \pm 0,6 | 11,0 \pm 0,6 | 5,9 \pm 0,5 | 100. | 8893 |
| May 78 | 19,5 \pm 0,8 | 9,6 \pm 0,6 | 14,1 \pm 0,7 | 52,0 \pm 1,0 | 3,2 \pm 0,3 | 1,6 \pm 0,2 | 100. | 72,8 \pm 0,9 | 11,4 \pm 0,7 | 10,3 \pm 0,6 | 5,5 \pm 0,5 | 100. | 9003 |

(SCB, Press release no. 1978:181, 9 June 1978)

The following papers are available in English :

Staffan Walhström and Staffan Söllander, "The Swedish Party Sympathy Surveys. A summary of the Preparatory Research Work", Statistik Tidskrift, 1974 : 2.

"Political Opinion Survey. The May 1975 Party Sympathy Survey", Statistical Report, Be 1975 : 5 (translated), National Central Bureau of Statistics.

"Party Preference Survey, May 1978", Press Release No. 1978 : 181 (translated), National Central Bureau of Statistics.

from Staffan SÖLLANDER
National Central Bureau of Statistics

HOUSEHOLD SURVEYS IN FRANCE : ON THE 1976 MASTER-SAMPLE

From the 1975 population census results, INSEE renewed the samples of its household surveys, as it usually does at each census. The topic of the paper is to present the sample used for most of those surveys (except for the Employment surveys based on an area sample), known as the "master-sample".

It results from a multi-stage random sampling, with stratification of the sampling units.

1. Primary sampling units

1.1 Definition - stratification

The Primary Sampling Units consist of the rural "communes" gathered by "canton" on the one hand, and the urban units on the other hand.

Those primary units were simultaneously stratified by geographic area and by category. The geographic areas consist of the eight ZEAT, resulting of the grouping of the 22 official "regions". The categories of primary units or strata are as follows :

- Stratum 0 : completely rural "cantons"
- Stratum 1 : rural fractions of partly urban "cantons"
- Stratum 2 : urban units under 20,000 inhabitants
- Stratum 3 : urban units of 20,000 to 100,000 inhabitants
- Stratum 4 : urban units of 100,000 to 2,000,000 inhabitants
- Stratum 5 : urban unit of Paris (except for Paris city)
- Stratum 6 : Paris city.

In its "a" columns, the following table gives the distribution of the 21 million censused dwelling units according to the regions and strata.

1.2 Sampling of the primary units

The number of units to be selected for each "ZEAT x stratum" cell was first fixed at the rate of one sample unit for 50 to 55,000 dwellings of the census. With this rate, for a 1/2000th survey, a rather frequent sampling rate at INSEE, the interviewer in charge of a sample primary unit, will have to visit 25 to 27 or 28 dwellings among those of the census (1/2000th of 50 to 55,000) plus some dwellings built after the Population Census, which appears to be quite a reasonable workload. However, this rule of "one sample primary unit for 50 to 55,000 dwellings" is actually applied only in the strata 0 to 3, where the size of the primary units is much smaller than those number. In stratum 4, made of urban units of 100,000 to 2,000,000 inhabitants, all the units have been automatically retained, each of them consisting of around 50,000 dwellings and very often much more. Of course, the urban unit of Paris (strata 5 and 6) was also automatically retained.

For strata 0 to 3, the numbers of sample primary units, fixed at the level of each ZEAT, have then been distributed among the "regions" of the ZEAT, as proportionally to the numbers of the dwellings of the census as the obligation to result in integers allows it. We took dispositions so that the over- and sub-representations are compensated between the strata 0 to 3 of each "region".

The numbers of sample primary units finally selected appear in the following table in its "b" columns. Let us point out that for the "1. Ile de France" and "3. Nord-Pas de Calais" ZEAT, the strata 0 and 1 were processed in bulk, strata 0 being too small to be processed separately.

The numbers of the sample primary units of strata 0 and 3 being fixed, they were drawn at the level of the "region x stratum" cell, (each primary unit having a probability of being selected proportional to its number of dwellings from the census).

Those sample primary units are invariable between two consecutive censuses (except for the replacement of some of the smallest of them when nearly half of the dwellings were visited), which accounts for the recruitment of an interviewer who lives on the spot and will carry out all the national surveys, together with the setting-up and updating of a "new dwellings" sampling frame (see § 3.2).

2. Secondary sampling units

The "communes" (1) of the sample primary units coming from the first sampling stage, make up the secondary sampling units.

In strata 0 and 1, where the primary units consist of rather similar rural "communes", some of these communes are selected with probabilities proportional to the numbers of the dwellings at the census (two communes for a 1/2000th survey, for example).

For each new survey, such a random selection is made ; each "commune" can thus be used in a variable number of surveys.

In strata 2 and plus, the primary units are urban units whose "communes" are more important and also more different. That is why they have been stratified in each unit, making a difference between :

- the most important "commune(s)" (most of the time, the only main town) that are automatically retained
- the other "communes" among which only some of them are selected (with probabilities proportional to the numbers of dwellings at the census).

In opposition to the rural "communes" of strata 0 and 1, the urban "communes" of strata 2 and plus, automatically retained or randomly selected are unvariable for all the surveys between two consecutive censuses (except for the replacement of some of the smallest ones when nearly half of the dwellings have been visited).

The urban unit of Paris is processed to the 2nd stage, like all the urban units : the City of Paris (stratum 6) and the 27 biggest outlying "communes" (stratum 5) are automatically kept, while among the other outlying "communes", 72 have been selected.

3. Tertiary sampling units

The dwellings of the "communes" selected after the first two sampling stages, make up the tertiary sampling units. Two cases must be considered depending on whether it deals with dwellings from the census or dwellings built after the population census.

3.1 Dwellings from the census

The number of the sample dwellings to draw from each "commune", among the dwellings of the census, depends on the sampling rate of the survey and on the sampling first and second stages. It is calculated by the "Sampling Methods" Division of the "Population-Household" Department.

The random selection of the sample dwellings is then made in each commune in a systematic way (equidistance) on a list including a variable fraction of the whole of the dwellings of the census of each "commune". For example, for all the rural "communes" and for the smallest urban "communes", the list includes 1 out of 2 dwellings of the 4/5th of the census. For the urban "communes" (apart from the smallest ones) it includes 1 out of k dwellings of the 1/5th of the population census, the number k changing from one "commune" to another and having been calculated so that the list allows to draw a set of surveys corresponding to the 1/20th rate.

The lists made from the 4/5th are manual and the selections are made manually. The lists made from the 1/5th have been realized by computer and the selections are made in the same way.

When the random selection is realized for the successive surveys, one takes care to eliminate the dwellings already selected for a preceding survey.

3.2 The dwellings built after the census

In each sample "commune", a sampling frame of "new dwellings" is made up as follows. From the monthly lists of building licences that the Regional Equipment Services send to the Regional Directions of INSEE, those select a random sample of licenced dwellings in the "communes" of the master-sample ; the rate of the sample change from one "commune" to another : from 1 out of 20 in the big cities that have been automatically retained when the 1 out of 1 secondary units are randomized, or the whole of the licenced dwellings, in the smallest "communes" and specially in the rural "communes".

This sampling of licenced dwellings started in the January 1972 lists (to take into account the building time) and is going on in monthly lists, as they are issued. Every six months, in June and December of each year, the Regional Directions of INSEE get information through all the appropriate means (consultation with the directions of Departments of Equipment, with local administrations, with builders, visit on the field ...) to know which dwellings are completed among the authorized and selected ones. When the completion of works is stated, an interviewer usually gets there to "identify" the dwelling : its location in the building, the name of the household living in it ... It is from the sampling frame, thus made and updated every six months, that the dwellings built after the census are drawn for each new survey.

(1) The "commune" is the smallest administrative division in France. There are about 36,000 "communes" in France.

4. The address-form (fiche-adresse)

is the document that allows the interviewer to trace back the selected dwelling ; among other things, it gives the address of the dwelling (street and number), the name of the people living in (whenever) the occupation of the head of the household,...

To fill in the address-form (fiche-adresse), the census questionnaires and the equivalent files of the recent dwellings have to be looked for manually. Indeed, in the census dwelling lists that are used for the random selection, the dwellings are described only by their identifying number useful for the Census (district, building and dwelling questionnaires number). Those identifying numbers allow to trace back the questionnaires corresponding to the selected dwellings and to clearly note down all the information reported in the address-form (fiche-adresse). The situation is similar for the dwellings built after the census.

So, we see that, in opposition to what happens for firm and factory samples, for which there are nominative files, the random drawing of a dwelling sample implies a great deal of hand work.

5. Time necessary to the setting-up of the master-sample

In February 1976, a year after the census, the "Commune" Data tape, allowed to set up sampling primary and secondary unit lists classified by region and stratum. The random selection of sample primary and secondary units has then been realized by the "Sampling Methods" Division and the list transmitted to Regional Directions in May-June 1976. It was then necessary to make up the "new dwellings" sampling frame in the sample "communes", enroll interviewers living close to the various units, and above all wait until the 1/5th file, that was supposed to be used for the random selection of the dwellings of the most important towns and "communes", is computerized ; which happened in February-March 1977. Is it sure that if a smaller segment, for example 1/20th, like for the 1968 Census, existed, it would allow a various months' gain.

STRUCTURE OF THE MASTER SAMPLE
Thousands of dwellings of the census (col.a)
and number of primary units sample (col.b)

| ZEAT Région | Strate | 0 | | 1 | | 2 | | 3 | | 4 (2) | | Ensemble | |
|---|---|------------------|-------|---------|-------|---------|-------|---------|---------|----------|---------|----------|---------|
| | | a | b | a | b | a | b | a | b | a | b | a | b |
| | | 1. ILE DE FRANCE | 21.2 | 1 | 136.1 | 2 | 155.2 | 3 | 105.4 | 2 | 3 680.1 | 3 | 4 098.0 |
| 21 Champagne-Ardenne | 117.0 | 2 | 87.3 | 2 | 80.7 | 1 | 100.8 | 2 | 121.9 | 2 | 507.6 | 9 | |
| 22 Picardie | 100.0 | 2 | 171.9 | 3 | 132.5 | 3 | 164.9 | 3 | 54.5 | 1 | 623.8 | 12 | |
| 23 Haute-Normandie | 62.7 | 1 | 119.8 | 2 | 91.9 | 2 | 79.0 | 2 | 234.9 | 2 | 588.2 | 9 | |
| 24 Centre | 160.0 | 3 | 224.4 | 4 | 173.9 | 4 | 157.7 | 3 | 168.8 | 2 | 884.8 | 16 | |
| 25 Basse-Normandie | 155.3 | 3 | 110.2 | 2 | 125.7 | 2 | 70.9 | 2 | 61.9 | 1 | 524.1 | 10 | |
| 26 Bourgogne | 184.4 | 4 | 147.5 | 3 | 127.1 | 2 | 138.8 | 2 | 76.6 | 1 | 674.4 | 12 | |
| 2. BASSIN PARISIEN | 779.4 | 15 | 861.0 | 16 | 731.8 | 14 | 712.1 | 14 | 718.5 | 9 | 3 802.9 | 68 | |
| 3. NORD - PAS DE CALAIS | 30.6 | 0 | 151.7 | 3 | 175.5 | 3 | 142.1 | 3 | 841.4 | 9 | 1 341.3 | 18 | |
| 41 Lorraine | 79.7 | 1 | 155.4 | 3 | 181.1 | 3 | 144.0 | 3 | 256.0 | 4 | 817.1 | 14 | |
| 42 Alsace | 23.9 | 1 | 108.6 | 2 | 112.8 | 2 | 71.4 | 1 | 220.9 | 2 | 537.6 | 8 | |
| 43 Franche-Comté | 98.2 | 2 | 78.4 | 1 | 79.7 | 2 | 59.3 | 1 | 87.5 | 2 | 403.1 | 8 | |
| 4. NORD-EST | 201.8 | 4 | 342.4 | 6 | 374.6 | 7 | 274.7 | 5 | 564.3 | 8 | 1 757.7 | 30 | |
| 52 Pays de la Loire | 211.4 | 4 | 224.7 | 4 | 188.8 | 4 | 83.0 | 2 | 356.8 | 4 | 1 064.7 | 18 | |
| 53 Bretagne | 226.3 | 4 | 265.9 | 5 | 228.1 | 5 | 112.2 | 2 | 191.7 | 3 | 1 024.2 | 19 | |
| 54 Poitou-Charentes | 186.5 | 4 | 130.4 | 3 | 88.6 | 1 | 129.7 | 2 | 74.3 | 2 | 609.6 | 12 | |
| 5. OUEST | 624.2 | 12 | 621.0 | 12 | 505.4 | 10 | 325.0 | 6 | 622.8 | 9 | 2 698.4 | 49 | |
| 72 Aquitaine | 209.1 | 4 | 174.4 | 3 | 168.4 | 3 | 133.5 | 3 | 333.0 | 3 | 1 018.4 | 16 | |
| 73 Midi-Pyrénées | 261.0 | 5 | 144.4 | 3 | 158.4 | 3 | 153.0 | 3 | 193.1 | 1 | 909.9 | 15 | |
| 74 Limousin | 133.3 | 3 | 48.9 | 1 | 44.7 | 1 | 31.9 | 0 | 67.2 | 1 | 326.1 | 6 | |
| 7. SUD-OUEST | 603.3 | 12 | 367.8 | 7 | 371.5 | 7 | 318.4 | 6 | 593.4 | 5 | 2 254.4 | 37 | |
| 82 Rhône-Alpes | 214.1 | 4 | 345.3 | 7 | 278.6 | 5 | 334.6 | 6 | 797.1 | 5 | 1 969.7 | 27 | |
| 83 Auvergne | 169.6 | 3 | 103.8 | 2 | 96.4 | 2 | 109.3 | 2 | 98.4 | 1 | 577.6 | 10 | |
| 8. CENTRE-EST | 383.8 | 7 | 449.1 | 9 | 374.9 | 7 | 443.9 | 8 | 895.6 | 6 | 2 547.3 | 37 | |
| 91 Languedoc-Roussillon | 122.8 | 2 | 156.7 | 3 | 227.3 | 5 | 121.3 | 2 | 186.5 | 3 | 814.7 | 15 | |
| 92 Provence - Alpes - Côte d'Azur - Corse | 185.9 | 4 | 118.3 | 2 | 185.4 | 3 | 239.6 | 5 | 1 035.7 | 6 | 1 765.0 | 20 | |
| 9. MEDITERRANEE | 308.7 | 6 | 274.9 | 5 | 412.8 | 8 | 360.9 | 7 | 1 223.3 | 9 | 2 579.6 | 35 | |
| TOTAL FRANCE | Logements (10 ³) | 2 952.9 | | 3 204.0 | | 3 101.8 | | 2 682.5 | | 9 138.5 | | 21 079.6 | |
| | UP - Echantillon | 57 | | 60 | | 59 | | 51 | | 58 | | 285 | |
| | Logements (10 ³)/UP - Echantil. | 51.8 | | 53.4 | | 52.6 | | 52.6 | | 157.6(3) | | 95.8(4) | |

(1) numbers rounded to 1/10th of the next thousand (3) average of the 58 urban units
 (2) including strata 5 & 6 (urban units of Paris) (4) average of the 57 urban units other than those of Paris

CORRELATION RATIO AND SAMPLING DESIGN

The purpose of this paper is to provide a rule of rapid estimation of the size of a sample when the sampling design belongs to the following types :

- Representative stratified sample (SR)
- Equal cluster sampling (G)
- Two stage sampling (2D) :
 - . Selection of primaries with probabilities proportional to size
 - . Selection of a constant number of secondary units per primary sampling units.

These three - very classical - designs are, at least in France, the most frequently used in market researchs and public opinion surveys.

1. Notations used

We try to estimate the mean \bar{Y} of a given value y in the universe, and whose variance is V . Besides, information is available (for instance in the form of tables) on the y distribution according to a criterion x (x being either quantitative, qualitative ordinal or nominal).

Under these conditions, the total variance can be analysed as : $V = B + W$ where B is the variance of conditional means (interclass variance) and W the mean of conditional variance (intraclass variance).

Finally, the intensity of linkage between the 2 variables y and x is measured by the correlation ratio :

$$\eta^2_{y|x} = \eta^2 = \frac{B}{V}$$

with $0 \leq \eta^2 \leq 1$. Usually, the correlation ratio is not known, but it can be estimated from a two-entries table.

It is assumed that the desired accuracy is measured by the sampling variance obtained from a selection according to the method M , E_M .

2. Sampling variances

With these notations, for a sample of size t , we get :

- simple random sampling (e) : $E_e = \frac{V}{t}$
- representative stratified sample : $E_{SR} = \frac{W}{t}$
- cluster sampling : $E_G = \frac{BN_0}{t}$

with N_0 : size of the cluster,
 m selected clusters : $t = m N_0$

- two-stage sampling : $E_{2D} = \frac{B n_0}{t} + \frac{W}{t}$

with : n_0 number of secondary sampling units per primary sampling unit,
 m selected primary sampling units : $t = m n_0$.

With these four relations, we can calculate the design effect, in relation to the simple random sampling.

If we call $d(M)$ the effect of design M in relation to the simple random sampling of equal size, we get :

$$d(SR) = 1 - \eta^2 ; d(G) = N_0 \eta^2 ; d(2D) = (1 - \eta^2) + n_0 \eta^2$$

with : $d(M) = E_M / E_e$

The first relation is well known ; to the contrary, we have no trace of any publication of the last two.

3. Application

It will be easy to verify, for these three methods, that if t is the sample size obtained by simple random sampling, and t_M , the size of the sample obtained through method M and providing the same accuracy :

$$t_M = d(M) t$$

When we have some notion of the value taken by V and that of η^2 , the size of the required sample can be estimated very rapidly.

4. Complementary remarks

Three remarks can be made :

1) When the total survey budget is fixed in advance, these relations make it possible to determine the sampling method and the optimum size sample.

Cluster sampling : if we call c_m , the unit cost of the interview for method M , we usually get :

$$\frac{c_m}{c_e} = K > 1$$

For a given budget, we can achieve $t_e = m N_0$ interviews by the simple random sampling method and $t_m = K m N_0$ by cluster sampling.

In this case :
$$\frac{E_e(t_e)}{E_e(t_m)} = \gamma^2 \frac{N_0}{K}$$

If the cluster is the household (in France $\bar{N}_0 = 2,88$ in the National Population Census of 1975) : since γ^2 is at most equal to 1, the cluster sampling is more efficient than the simple random sampling method, for a given budget, if the cost ratio K is superior to 2,88... which, in France, corresponds rather well to the cost of surveys carried out by private institutes.

2) From a pedagogical standpoint, this kind of approach allows beginners to understand rapidly the advantages and disadvantages of the various methods ; especially the opposition between stratification and clusters.

3) The comparison between various γ^2 for different control variables x_i , permits to choose the best related variable for a given division... Unfortunately, indeed, the value of γ^2 varies according to the division in classes of the adopted variable x ; despite that, this value stabilizes very rapidly.

It should be the task of national official organizations in charge of publishing the results of surveys or censuses, to publish, in the same time, tables of γ^2 values for various variables x and y , and studies on the stability of γ^2 for a division according to x , and the stability throughout time.

M. DEROO
I.A.S.S., France

QUESTION/ANSWER

(EDITOR : L. KISH)

WHAT ARE THE METHODS WHICH HAVE BEEN USED IN THE ENUMERATION OF NOMADS OR SEMI-NOMADS IN CENSUSES OR SURVEYS IN AFRICA AND WHAT ARE THEIR ADVANTAGES AND DISADVANTAGES ? - ANSWER BY K.T. DE GRAFT-JOHNSON

This Question/Answer 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 because he expects to find experts for special topics.

Send your questions to the IASS Secretariat in Paris, c/o INSEE, 18 boulevard Adolphe-Finard, 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, USA. But if you indicate that you sent it only to one of those, a copy will be sent to the other.

This is a second question synthesized from several concerned with sampling nomads 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.

Question

What are the methods which have been used in the enumeration of nomads or semi-nomads in censuses or surveys in Africa and what are their advantages and disadvantages ?

Answer by K.T. de Graft-Johnson, UN/ECA, Addis Ababa, Ethiopia

Six methods have been used in Africa for the enumeration of nomads or semi-nomads in censuses or surveys. These are :

- (a) Group Assembly Method
- (b) Social Organization or Hierarchical Approach
- (c) The Enumeration or Geographical Area Approach
- (d) The Waterpoint Method

- (e) The Camp Approach
- (f) The Dual Reporting System

A seventh method, that of satellite imagery followed by field identification, has been suggested but used only for the enumeration of nomadic livestock.

Before discussing the merits and demerits of each of these methods, it is relevant to note that a census or survey aims not only at counting the population, but also at obtaining some basic information on their characteristics.

The following are the advantages and disadvantages of the six methods mentioned above :

(a) The Group Assembly Method : The approach is useful when the nomadic tribes are amenable to discipline and accept the authority of those summoning them to a central place. The method has been tried with some success in North Africa and Mali and is less costly for the census or survey organizers. However, in view of difficulties and cost of travel to the place of assembly, the response usually tends to be poor in the absence of sufficient incentives.

(b) The Hierarchical Approach : This method makes use of the social, usually feudal, structure of the nomads. Thus among a nomadic tribe, there would usually be a chief. Under him would be a sub-chief and under the sub-chief a sub-sub-chief and so on. Under the lowest level of the sub-chiefs would be the heads of households. Each such sub-sub-chief "maintains" a list of heads of households under him. By making use of this hierarchical structure, it is possible to contact these heads and enumerate their households in censuses, or in surveys with multistage designs.

When the objectives of the census or survey are properly understood and there is complete co-operation from the chief, this approach can lead to satisfactory results, provided the chiefs are trusted by the nomadic heads of households. However there are two main problems with this method : defects in the lists of household heads and the cost of tracing the heads of households in order to enumerate them.

(c) The Enumeration Area (EA) Approach : This is the conventional approach for enumerating the settled population in a census. It can be extended to the nomadic or semi-nomadic population in those cases where the nomads roam within well-defined and reasonable territorial limits. In such cases, the territory can be divided into convenient geographical areas and canvassed by enumerators. But, if the nomads roam far and fast, in order to avoid omissions and duplications, the enumeration must be completed in a short time, preferably one day. This implies using a large force of enumerators. This is its main defect, since the manpower requirements for a day's enumeration of the nomads may be beyond the resources of the census or survey organization.

(d) The Waterhole Approach : The main reason nomads wander around is to search for water and grazing land. Thus if a complete frame of waterpoints in the area could be maintained and an enumerator stationed at each of them for the assumed maximum period during which a nomad will have to water his herd or flock a census of the nomadic population can be obtained from interviewing those who come to the waterholes. There are four defects of this method : Firstly, the list of waterpoints tends to be incomplete. Secondly, some nomads may avoid waterpoints where enumerators are stationed. Thirdly, since watering patterns vary, the assumed maximum period for watering may not be the real maximum. Finally, herds may be brought to the waterpoints by persons who have little knowledge of the households.

(e) The Camp Approach : A list of nomadic camps could be prepared and their location shown. If a comprehensive list can be prepared, this method has the advantage of being the easiest as far as enumeration is concerned. However, the lists are usually incomplete and the descriptions of their locations in the semi-arid terrain are usually misleading. In addition, in some countries, these camps keep on being moved from one site to another.

(f) Dual Reporting System : This is actually not a distinct method but an independent combination of any two of the methods outlined above. By counting these nomads from independent sources and applying the Chandrasekaran-Deming formula, namely :

$$\hat{N} = \frac{\hat{N}_1 \hat{N}_2}{M}, \text{ where :}$$

\hat{N} = the dual estimate of nomads

\hat{N}_1 = the estimated number of nomads from one independent source

\hat{N}_2 = the estimated number of nomads from the other independent source

M = the number of nomads common to both sources

The dual reporting method suffers from errors of mismatching and non-matching, the latter of which could lead to inflated estimates of the number of nomads. But mismatching, through inflated M , can produce underestimates, as does the common tendency for positive correlations in the noncoverages with both methods.

It should be stressed that in theory at least there is no preferred method. The choice of an enumeration approach will depend on conditions in a particular country. In general pre-tests are necessary before making the final choice of a method.

A comprehensive paper, E/CN.14/CAS 10/16, by the Economic Commission for Africa, Addis Ababa entitled "Study on Special Techniques for Enumerating Nomads in African Censuses and Surveys", summarises the most important case studies and makes a number of recommendations.

Editor's note

With regard to methods c, d, and e, it is important to clarify some additional flexibility that can be introduced with randomization of the time of enumeration. It is true that nomads may wander between (c) enumeration

areas, (d) waterholes, and (e) camp sites. Therefore, if the enumeration is not done at one time (in one day ?) all over, there will be frame errors of omissions and of duplications. However these errors can be randomized to have zero expectations, though they introduce errors into censuses, or additional errors into samples. It is necessary for zero expectations of errors to randomize the times of enumeration with regard to presence and absence of the nomads at the selected sites. For example, the nomads should not know of the selection to be able to avoid (or to approach) the site. Duplicate appearances in the sample are to be accepted to maintain unbiased expectations.

In other words, each member of the population should be identifiable with one site (area) for each "point" of time - where the "point" is the time for enumeration at the site, say a day. However, their moves must not be influenced by the selections made. This can be a difficult requirement but not as difficult as requiring or assuming immobility over the enumeration of the entire sample- or of the census, because this can also be done with this flexible method.

NEWS FROM THE ASSOCIATION

VISIT OF PERSONALITIES TO PARIS

The IASS Secretariat in Paris received the visit of S.H. KHAMIS, a Council member, V.R. RAO and B. DAVIES during the 1978 fourth term.

PREPARATION OF THE MEETING PROGRAM OF THE 1981 ISI-IASS SESSION

It is necessary to already prepare the 1981 ISI-IASS Session. For that purpose, T. DALENIUS, president elect, asks IASS members to let know their suggestions on the topics they would like to see dealt with during that Session.

The suggestions should be mailed to Dr T. DALENIUS, Division of Applied Mathematics, Brown University, Providence, R.I. 02912, USA, and two copies forwarded to the IASS Secretariat in Paris.

BIBLIOGRAPHY

Group of African Demography (ISP, INED, INSEE, MICOOP, ORSTOM), Regional Institute for Population Studies . -

"Bulletin de liaison" - Special Newsletter n°13, 4th term of 1978, 125 p.

- Demographic Aspects of the urbanization, L.A. KOSINSKI
- Census data used by indirect estimation methods of demographic parameters (1980 censuses), K. HILL
- Historical estimations of the world population : an assessment, J.D. DURAND
- Dual record demographic surveys : a re-assessment, J.G. BLACKER
- New estimates of fertility and child mortality in Africa, South of the Sahara, O. ADEGBOLA
- Sampling technics for the World Fertility Survey, V.K. VERMA

Statistical Institute for Asia and the Pacific . -

"Summary Record of the Six-Week Seminar on Planning and Conduct of Sample Surveys with Emphasis on Error Control and Use of Survey Data; May 6-June 17, 1977", Tokyo, Japan ; Chief Organizer : Dr M.N. Murthy, May 1978, 76 p.

Statistical Services Field, Statistics Canada . -

"Survey Methodology" - Vol. 4, n°1, June 1978, 124 p.

- The Utilization of Administrative Records for Statistical Purposes, L.E. ROWEBOTTOM
- The Effect of a Two-Stage Sample Design on Tests of Independence, J. COWAN and D.A. BINDER
- Approximate Tests of Independence and Goodness of Fit Based on Stratified Multi-Stage Samples, I.P. FELLEGI
- A Survey Design System for the Measurement of Truck Cargo Flows in Peru, A. SATIN and R. RYAN
- A Comparison of Correlated Response Variance Estimates Obtained in the 1961, 1971 and 1976 Censuses, K.P. KROTKI and C.J. HILL
- A Study of Refusal Rates to the Physical Measures Component of the Canada Health Survey, B.N. CHINNAPPA and B. WILLS
- An Estimate of the Efficiency of Raking Ratio Estimators under Simple Random Sampling, M.D. BANKIER

International Statistical Institute . -

"World Fertility Survey Technical Bulletins" -

- Some notes on statistical problems likely to arise in the analysis of WFS surveys, Sir M. KENDALL, n°1/Tech. 441, October 1976, 18 p.
- Path analysis and model building, M.G. KENDALL and C.A. O'MUIRCHARTAIGH, n°2/Tech 414, March 1977, 27 p.

"Scientific Reports" -

- Assessing the availability of fertility regulation methods : report on a methodological study, G. RODRIGUEZ, n°1, February 1977, 56 p.
- Language problems in demographic field work in Africa : the case of the Cameroon fertility survey, H. WARE, n°2, October 1977, 48 p.

"World Fertility Survey Semi-Annual Report" - January 1 - June 30, 1978, 49 p.

Statistical Services Field, Statistics Canada . -

"Quarterly Bulletin of the Methodology Divisions", n°26, juillet 1978

- Longitudinal Estimation in LFS
- The Occupational Employment Survey
- Intercity Bus Passenger Origin and Destination Pilot Survey

D.H. Flaherty . -

"Final Report of the Bellagio Conference on Privacy, Confidentiality, and the Use of Government Microdata for Research and Statistical Purposes", Statistical Reporter, May 1978, n°78-8, p. 274-278.

W.T. Federer . -

"Some Remarks on Statistical Education", The American Statistician, November 1978, vol. 32, n°4, p. 117-121.

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 these subjects among statisticians, organisations, institutions, governments, and the general public in different countries of the world.

OFFICERS AND MEMBERS OF THE COUNCIL

President (1978-79)
President elect (1978-79)
Executive Director (acting) (1978-79)
Vice-presidents (1978-79)

Scientific Secretary (1978-79)
Treasurer (1978-79)
Director of the ISI Permanent Office
Members (1978-79)

(1978-81)

P.V. Sukhatme (India)
T. Dalenius (Sweden)
G. Théodore (France)
I. Fellegi (Canada)
L. Kish (USA)
A.J. Scott (New Zealand)
G. Théodore (France)
E. Lunenberg (Netherlands)

E. de Alba (Mexico)
F. Azorin (Spain)
S.L. Diop (Senegal)
J. Harewood (Trinidad)
S.H. Khamis (Lebanon)
M.N. Murthy (India)

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).

IASS REPRESENTATIVES

M. Akehossi (Niger)
O.C. Ajayi (Nigeria)
H. Al Shabarti (Iraq)
M. Athanassopoulos (Greece)
V. Balaban (Yugoslavia)
F.E. Betham (Western Samoa)
L. Biggeri (Italy)
L. Bodart (Belgium)
B. Boyes (USA)
S.L. Diop (Senegal)
H. Duong (Madagascar)
H. El Bashir (Sudan)

F. El Boustani (Syria)
I.P. Fellegi (Canada)
E.K. Foreman (Australia)
M. Fracalossi (Argentina)
K.T. de Graft-Johnson (Ethiopia)
J. Harewood (Trinidad-Tobago)
C. Koley (Papua-New Guinea)
A.A. Le Roux (Great-Britain)
R.N. Lodhia (Fiji Islands)
M. Maaroufi (Tunisia)
F. Madsen (Denmark)
H. Midzuno (Japan)

G. Nathan (Israel)
S. Nour Eldin (Egypt)
M. N'e Faustin (Gabon)
H.K. Oh (Thailand, Korea)
E. Rapaport (Sweden)
J.L. Sanchez-Crespo (Spain)
M. Simonpietri (Chad)
D. Singh (India)
H. Strecker (Fed. Rep. of Ger.)
G. Théodore (France)
I.B. Thomsen (Norway)
S. Tulya-Muhika (Uganda)
M.A. Wahab (Pakistan)

IASS CORRESPONDENTS

B.A. Bailar (USA)
E. Cansado (Chile)
E.K. Foreman (Australia)
M.R. Frankel (USA)
M. Gonzalez (USA)

K.T. de Graft-Johnson (Ethiopia)
C. Hayashi (Japan)
T.B. Jabine (USA)
Ms Kantorowitz (Israel)
M.N. Murthy (Japan)

Ranjan Kumar Som (USA)
J.L. Sanchez-Crespo (Spain)
M.P. Singh (Canada)
F. Madsen (Denmark, Finland, Norway)



