

## SURVEY STATISTICIAN N° 6

September 1981

## CONTENTS

	Page
1. Editor's letter	1
2. General information	
2.1 Advanced workshops in survey methods - survey sampling ; university of Kent, Canterbury, UK, September 22-25, 1981	2
2.2 Proceeding of the seminar on household surveys, Bangkok, September 15-26, 1980	2
2.3 Health surveys on health status and health care, March 9, 1981, Cairo, Egypt	3
2.4 IASS meeting, May 13, 1981	3
3. Papers	
3.1 Double sampling for estimation of domain means, O.P. Kathuria, S. Tewari	4
3.2 Subregional workshop on the development of an integrated statistical system for a developing country (with Uganda as special case), S. Tulya-Muhika, B. Kiregyera	7
3.3 Comparisons of multi-frame with single-frame sample designs using registration and voting survey data, E.T. Huang, C.R. Hogue, C.T. Isaki	8
3.4 On the problem of variance estimation for a deseasonalized series, K.M. Wolter, N.J. Monsour.	9
3.5 Redesign of the survey of retail sales (Australia), T. Karmel	10
3.6 Rectification announcement	11
4. News of the Association	
4.1 Obituary, by T. Dalenius	11
4.2 Buenos Aires Session, Argentina, November 30 - December 11, 1981	
4.2.1. Meetings organized jointly by IASS and ISI	11
4.2.2. Summary booklet	12
4.3 IASS program committee, 1983 Madrid session	12
4.4 New IASS correspondents	12
5. Bibliography	12

## 1. EDITOR'S LETTER

## ANNOUNCEMENT BY THE ISI PRESIDENT

ISI Newsletter, N°14, June 1981

Members of the statistical profession have become increasingly aware that Mr Carlos Noriega, former Director of the Argentine National Institute of Statistics, has disappeared following his arrest in Argentina in 1977. As a result, the ISI Bureau has received letters asking (a) what it has done to obtain information about the whereabouts of Mr Noriega ; (b) whether its 43rd session, scheduled to be held in Buenos Aires in November/December 1981, will be held if such information is not received from the Argentine authorities.

It is appropriate to explain the ISI position.

1. The disappearance of Carlos Noriega was brought to the attention of the ISI Bureau in the early days of the Manila session (December 1979). At that session an invitation to hold the 43rd session in Buenos Aires was expected to be extended to the ISI General Assembly at its December 12 meeting. (It may be recalled that the ISI Statutes stipulate that decisions on the seat of future sessions must be taken by the General Assembly at its ordinary sessions).

The Bureau decided that the General Assembly should be informed about the pending invitation and about the disappearance of Mr Noriega at its December 7 meeting. During that meeting lengthy discussion took place in which many members participated. The General Assembly decided that this discussion be considered "off the record".

Subsequently, the question was extensively and informally discussed by members in the corridors prior to the meeting of December 12. The Bureau informed the Ambassador of Argentina to the Philippines of the concern of the General Assembly about the disappearance of Mr Noriega.

In the December 12 meeting of the General Assembly three invitations for future sessions were extended : Argentina (1981), Spain (1983) and the Netherlands (1985). In order to facilitate a free discussion by the members, the diplomatic representatives of the inviting countries withdrew after having presented their invitations. The chairman stated that the Bureau recommended the acceptance of the three invitations. He invited comments but none were made. The chairmen then proposed the acceptance of the invitations. This proposal was adopted by acclamation without discussion.

In accordance with the statutes of ISI, the Bureau abides by the decision of the General Assembly that the 43rd session be held in Buenos Aires in 1981.

2. At the suggestion of a member, the General Assembly in Manila requested the Bureau to make enquiries about C. Noriega. The President of ISI has addressed such an enquiry to the Ambassador of Argentina to France. He has communicated the concern expressed by members, and has asked that information be provided about Mr Noriega. The Ambassador has promised to communicate his request to his Government and to keep the President informed.

E. MALINVAUD, President, ISI

## 2. GENERAL INFORMATION

### 2.1 ADVANCED WORKSHOPS IN SURVEY METHODS - SURVEY SAMPLING ; UNIVERSITY OF KENT, CANTERBURY, UK, 22-25 SEPTEMBER 1981

This is the first of a series of proposed workshops on different aspects of survey methods organised by the Survey Methods Centre of SCPR (Social & Community Planning Research).

The aim is to bring together 20 to 30 research workers, survey practitioners and lecturers with a specialist interest in the topic - in this case survey sampling.

The workshop will take the form of a series of lectures and discussions on sample design, sampling errors, sampling frames, non response, imputation, statistical matching and small domain estimation. The main sessions will be given by Professor Graham Kalton (Research Scientist at the Survey Research Center, ISR, Michigan and Professor of Biostatistics at Michigan University) and Denise Lievesley, a Principal Social Survey Officer at OPCS and head of Social Survey's Sampling Section.

The general content of the main sessions will be as follows :

- optimum sample designs
- sampling errors
- sampling frames
- non-response, imputation, etc.

#### Programme

The course will be divided into six half-day and two evening sessions.

For further information, please write to Jennifer Butler, Survey Methods Centre, Social and Community Planning Research, 35 Northampton Square, London EC1V 0AX, United Kingdom.

### 2.2 PROCEEDINGS OF THE SEMINAR ON HOUSEHOLD SURVEYS, BANGKOK, SEPTEMBER 15-26, 1980

The United Nations Economic and Social Commission for Asia and the Pacific organized a seminar on Household Surveys at Bangkok (Thailand) from 15 to 26 September 1980, with financial assistance from the Federal Republic of Germany. The seminar was attended by survey statisticians from 20 member and associate member countries of the region. The United Kingdom, the Netherlands and the Federal Republic of Germany also sent participants. Representatives of the United Nations Statistical Office, the United Nations Children's Fund, the United Nations Development Advisory Team for the Pacific, the Food and Agriculture Organization of the United Nations, the International Labour Organization, the World Health Organization, the World Bank and the Population Council also attended the Seminar.

The objective of the seminar was to review the current household survey programmes in the countries of the region, discuss the organizational and methodological problems of household surveys and the evaluation, analysis and use of survey data, and focus attention on the National Household Survey Capability Programme (NHSCP), sponsored by the United Nations in collaboration with other International agencies.

The review of current household survey programmes revealed the growing demand for statistical data based on household surveys for the formulation and evaluation of socio-economic policies, plans and programmes and outlined the recent developments in survey design methodology and organization. Subject-wise, survey efforts have been largely concentrated on labour force, employment and unemployment, incomes and expenditures, demographic characteristics, fertility, mortality and family planning. Nonetheless, there has been a growing interest in other subjects such as housing, nutrition, health and welfare, including the problems of women, children and youth. The importance of household surveys within the national statistical systems was stressed and the need for developing a permanent and integrated household survey system was emphasized.

In that context, the major role of the National Household Survey Capability Programme, sponsored by the United Nations with the support of the United Nations Development Programme, the United Nations Fund for Population activities, the United Nations Children's Fund and the World Bank for building up survey capability in developing countries was noted. The Programme, operated in collaboration with the specialized agencies and implemented with the assistance of the regional commissions of the United Nations aims at establishing in developing countries durable capability for the organization of a continuous and integrated survey programme to meet the statistical needs of socio-economic development. Recognizing the potential of and the need for such a programme in the region, the seminar enthusiastically welcomed and endorsed the NHSCP.

The seminar discussed the organizational problems of integrated household surveys and opined that integration could be contemplated in various forms, e.g. administrative integration, organizational integration, operational integration and conceptual integration. It could vary in degree from complete integration in all aspects at one end to mere co-ordination at the other. Administrative integration involved centralization of over-all responsibility for all large-scale household surveys in a single organization or, at least, effective co-ordination through an appropriate machinery for inter-departmental consultations, discussions and decisions. Organizational integration envisaged vertical integration down to the field level as well as horizontal integration of various functional agencies engaged in design mapping and frame preparation, recruitment and training of investigators, field operations and supervision and data processing. Operational integration could be total or partial, and covered planning and programming, design, field operations, subject coverage, processing and analysis. Conceptual integration aimed at uniformity and/or compatibility of the concepts, definitions and classifications used in different surveys. The seminar, therefore suggested that in the development of integrated household survey programmes, it would be necessary to define the nature, form and extent of integration needed, desirable and feasible within the constitutional, legal and administrative framework of the country, taking into account the financial, technical and technological constraints and the manpower and mechanical resources available.

The seminar considered the advantages and disadvantages of unisubject and multisubject surveys and noted that the latter naturally had all the advantages of integration and provided an opportunity to cross-tabulate and correlate information on the various aspects of household life ; but they had some disadvantages such as lengthy questionnaires, a multitude of concepts, definitions, code structures, procedures and methods of inquiry, long interviews, repeated visits, investigator fatigue, respondent resistance, complexities of training and supervision, compromises in design, concepts and definitions etc. On a balance of consideration, the seminar felt that an omnibus survey covering all sorts of subjects would be unwise and that a group of related subjects, related in conceptual and analytical terms, offered an optimum combination. Different combinations could be attempted in different rounds and the surveys so organized as to provide, over a period of time, an interlinked system. An alternative was a continuing survey on a focal subject with a

variable supplementary schedule. The choice of supplementary subjects for the different rounds could be so determined as to facilitate link-up between related subjects through repeated samples.

In discussing sample design for integrated household surveys, the seminar observed that an appropriate strategy would be to minimize the number of sampling stages and adopt small compact area units as the first stage sampling units in a possible two-stage sampling design. The first stage sample should preferably be large and the second stage sample variable in size depending on the subject under investigation. The seminar suggested a basic geographic stratification at the first stage of sampling to control geographic variation. It also discussed the advantages of selection of first stage sampling units with probability proportional to size, self-weighting designs and rotational sampling.

The seminar discussed in detail the conceptual, methodological and analytical problems of a wide range of household surveys, viz., demographic surveys, labour force surveys, income and expenditure surveys, food consumption and nutrition surveys, surveys of basic needs, time-budget surveys and surveys of household economic activities, both agricultural and non-agricultural. These matters called for closer attention by specialized groups. A subject of general concern, however, was the role of household surveys as a source of information on women, children and youth. This subject, especially so far as it concerned women, was given considerable attention in view of the feelings expressed that the role of women in household economic activity was not adequately captured in household surveys, that the requisite data concerning the status of women were not properly collected and that the household survey mechanism was generally biased against women. The seminar agreed that it was necessary to probe deeper into household life to detect age-sex disparities in the access to assets, employment, consumption, etc. The seminar also recognized that the work-patterns of women and children were different from those of men because of socio-cultural factors and that the concept of work and the related criteria for the classification of individuals as working and non-working used in employment surveys should be reviewed.

The seminar observed that time-disposition studies covering not only the time spent on economic activities but also on certain quasi-economic activities and household chores which prevented women from participating in economic activities would be useful. It was recognized in this connexion that while time-budget studies designed to ascertain the use of leisure as carried out in developed countries might not be relevant to the developing countries of the region, studies focused on identifying the economic pressures on men, women and children, especially in the poor and disadvantaged segments of the societies in developing countries might be useful. The seminar also discussed the desirability and possibilities of utilizing female investigators in household surveys to collect information from female respondents. The experience in this regard varied.

The seminar discussed methods of evaluation, analysis and uses of survey data and emphasized the need to gear data collection in household surveys to the anticipated needs of the users. It noted that household survey data were generally subject to sampling errors, biases and non-sampling errors, and stressed the need not only to control such errors and biases but also to evaluate them through resurveys, record checks, analysis of recall lapses, comparison with data available from other sources, etc. The use of interpenetrating subsamples and fractile graphical analysis was also suggested.

The seminar recognized the immense analytical potentialities of the demographic, social and socio-economic data collected through household surveys and the possible uses of such data in planning, national accounts and economic analysis and emphasized the need for effective consultation and co-operation between the producers and users of statistics.

The seminar underlined the need for sharing experiences in survey methodology and practice not only through seminars, workshops and conferences but also through documentation. It also stressed the need for training and technical assistance.

In conclusion, the seminar urged the national statistical offices to raise the status of statistical activity by involving policy-makers and gaining their support for building a strong statistical organization. It also underscored the need for external technical and financial assistance and drew attention to the possibilities of technical co-operation between developing countries.

## 2.3 HEALTH SURVEYS ON HEALTH STATUS AND HEALTH CARE, MARCH 9, 1981, CAIRO, EGYPT

A program was recently held in Cairo, Egypt, entitled Health Surveys on Health Status and Health Care. The program was part of the International congress on Statistics, Computer Science, and Social Research, and co-sponsored by the International Association of Survey Statisticians.

The session was held on Monday, March 9 at the school of Public Health, Ain Shams, University. The papers and speakers included :

1. Medical Care Expenditure Surveys : Methodological Issues, by Daniel G. Horvitz, Vice-President for Statistical Research (who presented the paper) and Ralph E. Folsom, Jr., Research Triangle Institute, Research Triangle Park, North Carolina.
2. The Health Profile of Egypt. Ramsis A. Gomaa, First Undersecretary of State, Ministry of Health, ARE, Cairo, Egypt.
3. The Present and Future Roles of National Health Surveys with Special Emphasis on Health Interview Surveys. Bernard G. Greenberg, Dean, School of Public Health, University of North Carolina, Chapel Hill, North Carolina.
4. Discussant - Robert A. Israel, Deputy Director, National Center for Health Statistics, Hyattsville, Maryland.
5. Chairman and Organizer - E. Earl Bryant, Associate Director National Center for Health Statistics, Hyattsville, Md.

## 2.4 IASS MEETING, MAY 13, 1981

On May 13, 1981, a meeting was held at the IASS Secretariat, in Paris, which was attended by : Lunenberg, Director of the ISI Permanent Office, Latter, IASS representatives : Sanchez-Crespo, Chairman of the 1981 Program Committee, O'Muircheartaigh, Chairman of the 1983 Program Committee, Théodore, Executive Director, Damiani, Hossenlopp.

Mr Sanchez-Crespo gave information regarding the meetings of the December 1981 session in Buenos Aires. Mr O'Muircheartaigh indicated the composition of the Program Committee for the 1983 session. He has asked IASS members to state their wishes and proposals to the Program Committee. The Committee intends to prepare, before the Buenos Aires Session, a list of 10 to 16 topics for the Madrid Session.

Mr Théodore announced that, during the Buenos Aires Session, Officers and members of the IASS Council will discuss the opportunity to issue, every year, a book containing the papers sent by members of the Association and not published in the Survey Statistician.

A meeting on migrations, organized by the ISI and IUSSP will be held in The Hague on September 8-10, 1981. It has been decided that IASS will be represented in this meeting by Pr Clairin, from INSEE.

A seminar on sampling surveys and censuses, intended for French-speaking statisticians, will be organized by the IASS, in Dakar, Senegal, in November 1981. Mr Scott from the World Fertility Survey, as well as Spanish delegates, have been asked to attend this seminar.

A visit to UNESCO was organized the same day, in view of having the IASS approved as a non-governmental organization, category C.

### 3. PAPERS

#### 3.1 DOUBLE SAMPLING FOR ESTIMATION OF DOMAIN MEANS

O.P. KATHURIA, S. TEWARI, Indian Agricultural Statistics Research Institute, New Delhi, India

##### INTRODUCTION

In a sample survey often the objective may be not only to estimate the population parameter such as mean, total or proportion for the character under study but also to obtain estimates for each of a number of classes into which the population may be subdivided. Such parts of a population constitute the domains of study or sub-populations. For instance, farmers may tend to grow different vegetables in parts of the same field, or grow different varieties of (say) wheat crop in the same season or in the next season. While the total area under the crop may not vary much from year to year, that under different varieties may change substantially as new and more promising varieties are introduced in the cropping system.

The problem of estimation of parameters of sub-populations was first pointed out by U.N. Sub-commission in 1950. Yates (1952, 1960) considered the problem of estimation of domain totals, means and proportions in case of simple random sampling. Durbin (1958), Hartley (1959) and Kish (1961) studied the problem of estimation of domain means within the frame-work of stratified random sampling. Scott & Smith (1971) examined in details various ways in which prior knowledge about the domain size could be incorporated into the estimates of means and totals. The present study aims at estimating the means of sub-populations, through double sampling approach. When sampling is done for two occasions with partial matching of units on the second occasion, estimates of the sub-population mean on the second occasion have also been obtained.

##### ESTIMATION OF SUB-POPULATION MEANS IN DOUBLE SAMPLING

###### A. Sampling Scheme

Consider a population of size  $N$  of which  $N_i$  units belong to the  $i^{th}$  domain. Our aim is to estimate the population mean  $\bar{y}_{N_i}$  of  $y$  based on  $N_i$  units ( $i = 1, 2, \dots, k$ ). Let  $x$  be a character highly correlated with  $y$  and  $\bar{x}_{N_i}$  the unknown mean of  $X$  for the  $i^{th}$  domain. Select a preliminary sample of size  $n'$  with SRSWOR on which the auxiliary character  $X$  alone is observed. A subsample of  $n$  units is taken out of  $n'$  units on which both  $X$  and  $Y$  are observed. Let  $n'_i$  and  $n_i$  be the number of units out of  $n'$  and  $n$  respectively which belong to the  $i^{th}$  domain. Obviously  $n'_i$  and  $n_i$  are both random variables. We assume that  $n'$  and  $n$  are sufficiently large such that all the domains of the population are represented in these samples.

###### B. Estimator and its variance

For estimating  $\bar{y}_{N_i}$ , the population mean of  $Y$  for the  $i^{th}$  domain, we propose the estimator

$$\bar{y}_{di} = \bar{y}_{n_i} + b_i (\bar{x}_{n'_i} - \bar{x}_{n_i}) \quad (1)$$

where  $\bar{y}_{n_i}$  is the sample mean of  $y$ , for the  $i^{th}$  domain based on  $n_i$  units and  $\bar{x}_{n_i}$  and  $\bar{x}_{n'_i}$  are the sample means of  $X$  for the  $i^{th}$  domain based on  $n_i$  and  $n'_i$  units respectively and  $b_i$  is the least square regression coefficient of  $Y$  on  $X$  in the  $i^{th}$  domain computed on  $n_i$  observations. Obviously,  $\bar{y}_{di}$  is biased, the bias in  $\bar{y}_{di}$  being given by

$$E(\bar{y}_{di}) - \bar{y}_{N_i} = \beta_i NN_i^{-1} (n_i^{-1} - n'^{-1}) [1 + (N - N_i) N_i^{-1} (n_i^{-1} + n'^{-1})] [\mu_{30}^{(i)} S_{x_i}^{-2} - \mu_{21}^{(i)} S_{xy_i}^{-1}]$$

(derivation of bias given in appendix I). If  $x$  and  $y$  follow bivariate normal distribution in each domain, the bias becomes approximately zero and accordingly the variance of  $\bar{y}_{di}$  may be obtained as

$$V(\bar{y}_{di}) = NN_i^{-1} [(n_i^{-1} - N^{-1}) \{n_i^{-1} - n_i^{-1}\} e_i^2] S_{y_i}^2 \quad (2)$$

where  $\rho_i$  is the population correlation coefficient between  $y$  and  $x$  and  $S_{y_i}^2$  is the mean sum of square of  $y$  in the  $i^{th}$  domain.

##### DOMAINS OF STUDY ON SECOND OCCASION

###### A. Sampling scheme and notations

To estimate the population mean of a character under study in the  $i^{th}$  domain on the second occasion, we shall consider the sampling scheme as given by Kuldorff (1963).

On the first occasion, a sample  $S_{1n}$  of size  $n$  is drawn with equal probability without replacement from the entire population. Let  $n_{1i}$  be the number of units in  $S_{1n}$  coming from  $i^{th}$  domain. On the second occasion, a sub-sample  $S_{2m}$  of size  $m$  units is drawn from  $S_{1n}$  with equal probability without replacement and a fresh sample  $S_{2\ell}$  of size  $\ell (= n, m)$  is selected independently with equal probability with replacement from those  $(N - n)$  units in the population which were not sampled on the first occasion. Let  $m_{2i}$  and  $\ell_i$  (both random variables) be the number of units in  $S_{2m}$  and  $S_{2\ell}$  respectively which belong to the  $i^{th}$  domain on the second occasion such that  $\sum_{i=1}^k m_{2i} = m$  and  $\sum_{i=1}^k \ell_i = \ell$  ( $i = 1, \dots, k$ ). Further let  $N_{\ell i}$  be the number of population units which belong to the  $i^{th}$  domain on  $\ell^{th}$  occasion ( $\ell = 1, 2$ ).

For simplicity, we assume that the population is sufficiently large such that  $N_{1i} = N_{2i} = N_i$ . Let

$y_{kij}$  : Value of the observation of the character under study for the  $j^{th}$  unit of the  $i^{th}$  domain on  $k^{th}$  occasion ( $k = 1, 2$ ),

$\bar{y}_{ki}$  : population mean of  $Y$  for the  $i$ th domain on  $k$ th occasion,

$S_{ki}^2$  : population mean sum of squares for the  $i$ th domain on  $k$ th occasion based on  $N_i$  units

$\rho_i$  : population correlation coefficient between the first and second occasion values on the units of the domain

Also, let the sample means based on matched units on first and second occasion, on the entire sample  $n_i$  on the first occasion and on unmatched units  $l_i$  on the second occasion in the  $i$ th domain be denoted by  $\bar{y}_{m_i}$ ,  $\bar{y}_{m_2}$ ,  $\bar{y}_{n_i}$  and  $\bar{y}_{l_i}$  respectively.

**B. Estimator and its variance**

We propose an estimator of  $\bar{y}_{2i}$  as

$$\hat{y}_{2i} = \phi_i \bar{y}_{m_i} + (1 - \phi_i) \bar{y}_{l_i} \quad (3)$$

where  $\phi_i$  is a suitably chosen constant for the  $i$ th domain such that  $V(\hat{y}_{2i})$  is minimum and  $\bar{y}_{m_i}$  is the regression estimator based on matched units on first and second occasion in the  $i$ th domain and is given by

$$\bar{y}_{m_i} = \bar{y}_{m_2i} + b_i (\bar{y}_{m_1i} - \bar{y}_{m_2i}) \quad (4)$$

Variance of  $\hat{y}_{2i}$  may be seen to be given by

$$V(\hat{y}_{2i}) = NN_i^{-1} n^{-1} (1 - \rho_i^2 e_i^2)^{-1} (1 - \rho_i^2 n^{-2} e_i^2)^{-1} S_{2i}^2 \quad (5)$$

(Derivation given in Appendix II)

Optimum value of  $l_i n^{-1}$  (=  $\mu$  say) at which the variance is least is

$$\mu = (1 + \sqrt{1 - \rho_i^2})^{-1} \quad \text{and optimum variance with } \mu \text{ is}$$

$$V_{opt}(\hat{y}_{2i}) = \frac{1}{2} NN_i^{-1} n^{-1} (1 + \sqrt{1 - \rho_i^2}) S_{2i}^2 \quad (6)$$

**ILLUSTRATION**

The theory developed in the preceding sections may be illustrated with the help of data of the sample survey for methodological investigations into high yielding varieties programme conducted by the I.A.S.R.I., New Delhi, India.

Treating the extent of the area under different high yielding/improved varieties as various domains of study, the data available from the area Estimation Enquiry and the Yield Estimation survey conducted during 1975-76 on wheat crop in Amritsar, Bullundur and Patiala districts of the Punjab state were utilized for the study on double sampling. Similarly, the data available from Area Enquiry conducted during 1975-76 and 1976-77 in Amritsar district of the Punjab state were used for estimating area under different high yielding/improved varieties during 1976-77.

The approach followed in the two studies is briefly described as follows :

**(i) Double sampling for estimation of sub-populations means**

The data of area of fields under different high yielding varieties of wheat crop for 204 cultivators in three districts viz., Amritsar, Jullundur and Patiala were treated as the population for the present study. From this population a preliminary sample of size  $n' = 100$ , was selected with equal probability without replacement and areas under different varieties were recorded for one field per cultivator. Further, a simple random sample of size  $n = 50$ , was drawn from  $n'$  units for estimation of yields of different varieties. The three main varieties of wheat grown in these three districts were W.G. - 357, HD-2009 and Kalyan Sona accounting for most of the area under high yielding varieties of wheat. All other varieties such as W.G. 377, PU-18, Sonalika, etc., which were sown in relatively smaller areas were grouped as "others". Apparently, the variability in the area and the yield of varieties in this group will be large. Using equations (1) and (2) for the estimator and its variance respective, the estimates average yield with standard error of the estimate for each of these varieties along with their corresponding average yield for the population of 204 cultivators area given in Table 1.

Estimates of average yield along with their percentage standard errors for each of the domains (varieties) under study (table 1)

Domain variety	Size of the domain $N_i$	Sample size $N_i$	$n$	Estimated average yield (Q/ha)	% S.E.	Average yield in the population (Q/ha)
W.G.-357	68	36	22	27.17	4.7	27.78
HD-2009	61	25	12	29.34	9.3	26.41
Kalyan Sona	41	21	7	21.41	11.5	27.96
"Others"	34	18	9	30.50	7.3	23.53

As it is clear from the table the estimated average yields on the basis of sample of 50 cultivators are broadly in agreement with their corresponding population average yields.

**(ii) Estimation of domain means on two occasions**

The data of 564 cultivators belonging to Amritsar district of the Punjab state for the Area Estimation Enquiry were treated as the population for the study. A simple random sample of 100 cultivators was drawn from this population during the year 1975-76 and the area under different high yielding varieties of wheat was recorded. About 50 per cent of the sampling units were retained in 1976-77 and remaining 50 per cent were drawn afresh from the rest of the population excluding the sample drawn during 1975-76. Using equations (3) and (5) the estimates of area under high yielding varieties of the crop and their percentage standard error along with the average area under different varieties in the population are given table 2. The estimated average area of each of these varieties without using the data of 1975-76 are given in column N° (7).

Estimation of average area along with their percentage standard error's for each of the domains (varieties) under study (table 2)

Domain (variety)	Size of the domain $N_i$	Sample size		Estimated average area (ha)	Present S.E.	Estimated average area without using the data of 1975-76 (ha)	
		$n_i$	$m_i$				
1	2	3	4	5	6	7	8
Kalyan Sona	147	26	17	0.92	5.02	0.90	0.94
W.G. 357	337	58	28	1.25	6.66	1.51	1.34
"Others"	80	15	5	1.25	10.35	1.01	1.35

From the table it clearly follows that the estimated average area under each of these varieties using the data of both the years is in agreement with corresponding population values. However, when data of 1975-76 have not been taken into account and the estimates are based on data of 1975-76 only, the estimated average area under each of the varieties (as shown in column 7) differs considerably from the corresponding population values.

REFERENCES

Cochran, W.G. (1963) : "Sampling Techniques" John Wiley and Sons, Inc., New York, 2nd edition  
 Durbin, J. (1958) : "Sampling theory for estimates based on fewer individuals than the number selected". Bull. Int. Stat. Inst. 36, 3, 113-119  
 Hartley, M.D. (1959) : "Analytic studies of survey data" Inst. di Statistica, Rome, Vol. in honour of Corrado Gini  
 Kuldorff, G. (1963) : "Some problems of optimum allocation for sampling on two successive occasions". Rev. Int. Stat. Inst. 31, 24-57  
 Kish, L. (1961) : "Design and estimation for subclasses, comparisons and analytical statistics. New Developments in Survey Sampling" (edited by Norman L. Johansen & Harry Smith, Jr) John Wiley and Sons, inc. 416-438  
 Patterson, H.C. (1950) : "Sampling on successive occasions with partial replacement of units" J.R.S.S., B, 12  
 Scott, A. and Smith, T.M.F. (1971) : "Bayes estimates for sub-classes in stratified sampling", J.A.S.A., 66, 834-836  
 Sukhatme, P.V. & Sukhatme B.V. (1970) : "Sampling theory of surveys with applications", Second Ed., page 195  
 U.N. Statistical Office (1950) : "The preparation of sampling survey reports" Paper series n°1.

APPENDIX I

Sukhatme & Sukhatme (1970) have obtained the expression for bias of the usual regression estimator  $\bar{y}_e - \bar{y}_n + \beta(\bar{x}_n - \bar{x}_n)$  as

$$E(\bar{y}_e) - \bar{y}_n = \left(\frac{1}{n} - \frac{1}{N}\right) \beta [M_{20} S_x^2 - M_{21} S_{xy}]$$

where  $M_{jk} = E[(x - \bar{x}_n)^j (y - \bar{y}_n)^k]$

Using this result, the bias in  $\bar{y}_{di}$  may be written as

$$E(\bar{y}_{di}) - \bar{y}_{ni} = E(n_i^{-1} - n_i'^{-1}) \beta_i [M_{20}^{(i)} S_{x_i}^2 - M_{21}^{(i)} S_{x_i y_i}]$$

where  $\beta_i$  is the population regression coefficient in the  $i^{th}$  domain and  $M_{jk}^{(i)} = E[(x_i - \bar{x}_{ni})^j (y_i - \bar{y}_{ni})^k]$

and  $S_{x_i}^2$  and  $S_{x_i y_i}$  are the mean sum of squares and sum of products of  $x$  and  $y$  in the  $i^{th}$  domain of the population.

It may be shown that  $E(n_i'^{-1}) = NN_i^{-1} n_i'^{-1} + NN_i^{-2} n_i'^{-2} (N - N_i)$

(a)

with a similar expression for  $E(n_i^{-1})$ . Therefore,

$$E(n_i'^{-1} - n_i^{-1}) = NN_i^{-1} [n_i^{-1} + N_i^{-2} (N - N_i) - n_i'^{-1} - N_i^{-2} (N - N_i)]$$

$$= NN_i^{-1} (n_i^{-1} - n_i'^{-1}) [1 + N_i (N - N_i) (n_i^{-1} + n_i'^{-1})]$$

APPENDIX II: Derivation of  $V(\hat{y}_{2i})$

$$V(\hat{y}_{2i}) = \phi_i^2 V(\bar{y}_{mi}) + (1 - \phi_i^2) V(\bar{y}_{ei}) + 2\phi_i(1 - \phi_i) Cov(\bar{y}_{mi}, \bar{y}_{ei}) \dots (b)$$

$\bar{y}_{mi}$  is of the same form as  $\bar{y}_{di}$  in equation (1). Hence replacing  $n'$  by  $n$ ,  $n$  by  $m$  and neglecting  $N^{-1}$ , and also noting that  $\rho_i$  is now the correlation between population units of the  $i^{th}$  domain on the first and second occasions, we get

$$V(\bar{y}_{mi}) = NN_i^{-1} (m^{-1} - \rho_i m^{-1} n^{-1} n_i^{-1}) S_{2i}^2$$

$$V(\bar{y}_{ei}) = E[V(\bar{y}_{ei} | \rho_i)] + V[E(\bar{y}_{ei} | \rho_i)]$$

The second term on the right hand side is zero, and  $V(\bar{y}_{ei} | \rho_i) = \left(\frac{1}{n} - \frac{1}{N_i}\right) S_{2i}^2$

Therefore,  $E V(\bar{y}_{ei} | \rho_i) = [E(\rho_i^{-1}) - N_i^{-1}] S_{2i}^2$

Using  $\ell_i$  and  $\ell$  in place of  $n_i'$  and  $n'$  equation (a) of Appendix I, we get

$$\begin{aligned} \{ \ell_i^{-1} \} - N_i^{-1} &= N N_i^{-1} \ell^{-1} [1 - \ell N^{-1} + \ell^{-1} N_i^{-1} (N - N_i)] \\ &= N N_i^{-1} \ell^{-1} \end{aligned}$$

to the first degree of approximation.

Therefore  $V(\bar{y}_{2i}) = N N_i^{-1} \ell^{-1} S_{2i}^2$

Similarly, we can also write

$$V(\bar{y}_{m_i}) = N N_i^{-1} m^{-1} S_{2i}^2 \quad \text{and} \quad V(\bar{y}_{n_i}) = N N_i^{-1} n^{-1} S_{2i}^2$$

Regarding  $\text{cov}(\bar{y}_{m_i}, \bar{y}_{2i})$ , it may be observed that for the sample mean  $\bar{y}_{n,2i}$  for the  $i^{\text{th}}$  domain on the second occasion

$$V(\bar{y}_{n,2i}) = \frac{1}{n} V(m\bar{y}_{m_i} + \ell\bar{y}_{2i})$$

i.e.  $N N_i^{-1} n^{-1} S_{2i}^2 = m^2 n^{-2} V(\bar{y}_{m_i}) + \ell^2 n^{-2} V(\bar{y}_{2i}) + 2m\ell n^{-2} \text{cov}(\bar{y}_{m_i}, \bar{y}_{2i})$

Therefore using the variance expression derived above, we get on simplification that  $\text{cov}(\bar{y}_{m_i}, \bar{y}_{2i}) = 0$

Therefore  $V(\hat{y}_{2i}) = \phi_i^2 N N_i^{-1} (m^{-1} - \ell m^{-1} n^{-1} \ell^2) S_{2i}^2 + (1 - \phi_i^2) N N_i^{-1} \ell^{-1} S_{2i}^2$

Minimising according to  $\phi_i$  and substituting in (b) we get the desired value of  $V(\hat{y}_{2i})$ .

### 3.2 SUBREGIONAL WORKSHOP ON THE DEVELOPMENT OF AN INTEGRATED STATISTICAL SYSTEM FOR A DEVELOPING COUNTRY (WITH UGANDA AS SPECIAL CASE)

S. TULYA -MUHIKA, B. KIREGYERA, Institute of Statistics and Applied Economics, Makerere University, Kampala, Uganda

#### A. PURPOSE

A workshop on the Development of an Integrated Statistical System for a Developing Country (with Uganda as a special case) was held at the Regional Institute of Statistics and Applied Economics, Makerere University, from 16 to 20 June 1980. It was sponsored by the United Nations Economic Commission for Africa, Statistics Division of the Ministry of Planning and Economic Development, Uganda and the Coffee Marketing Board, Uganda. The participants who numbered 47, came from Kenya, Tanzania, Zambia and Uganda.

Uganda was chosen as a special case because the once thriving statistical infrastructure in Uganda was utterly destroyed, like everything else during the reign of terror of Idi Amin. This is apparent in many ways. For instance: the Central Statistics Division is very thinly manned and very ill-equipped; the last rural (important surveys) were carried out in the country in the late sixties; the last publication of a statistical abstract was in 1974. One can easily characterise the Seventies as a decade of data sparsity in Uganda, leaving yawning gaps in data.

The task before the workshop was immense: it involved identification of problem areas associated with developing an Integrated Statistical System in a developing country, Uganda in particular, and formulation of concrete practical proposals and recommendations to the Government of Uganda and other relevant agencies for the development and sustenance of such a system. It also involved identification of data needs with priorities, charting out ways for improvement of data collection and publishing census and survey reports quickly, etc.

#### B. INTEGRATED NATIONAL HOUSEHOLD SURVEY SYSTEM (INHSS)

A strong case was made for immediate action to develop National Household Surveys in Uganda in an integrated framework in keeping with the objectives of the National Household Survey Capability Programme (NHSCP). In this connection, strong recommendations were made for:

- (i) Government to make funds available as a matter of urgency to facilitate resumption of nation-wide agricultural surveys which had been interrupted by the advent of Idi Amin.
- (ii) Carrying out follow-up demographic surveys and a Fertility Survey in keeping with the objectives of the WFS
- (iii) Resumption of annual surveys of industrial production and extending their coverage to encompass the informal sector, viz. handicrafts and other small-scale and cottage industries
- (iv) Carrying out any other household surveys as part of the NHSCP to provide data which, among other things, would facilitate evolution of an integrated system of social indicators
- (v) Establishment of viable points of integration of the system, viz. National Sample (and Cartographic Unit), Permanent Field Organization, a basis of which already exists in the form of permanent field staff that previously were used to carry out annual nation-wide agricultural surveys, and Data Processing and Analysis Unit.

A more detailed paper on the development of an INHSS in the context of the African conditions is being prepared for the 1981 IASS session.

It was, however, felt that it would not be enough to establish an INHSS as this can be meaningfully viable only in the context of development of an overall Integrated National Statistical Information System (INSIS). Recommendation was also made to establish and develop an autonomous and self-accounting National Statistical Agency (NSA) in Uganda as a symbiosis of the System of Surveys, System of Censuses and System of Administrative Records.

### 3.3 COMPARISONS OF MULTI-FRAME WITH SINGLE-FRAME SAMPLE DESIGNS USING REGISTRATION AND VOTING SURVEY DATA

E.T. HUANG, C.R. HOGUE, C.T. ISAKI, US Bureau of the Census

In 1976 this study was proposed to compare the cost efficiency of a two-frame sample design to a single-frame sample design for estimating minority registration and voting rates. Data from the 1976 registration and voting survey (RAV) and from county registration lists were used to estimate costs, variances, and covariances in eight county jurisdictions. In this paper five models denoted A through E are considered in the comparison of several single-frame and two-frame approaches.

The 1976 Registration and Voting Survey utilized four mutually exhaustive frames to cover the target population. The four mutually exclusive frames used in RAV consisted of (1) households from the 1970 Census tapes, (2) group quarters from the 1970 Census tapes, (3) new construction from building permits issued since 1970, and (4) area segments of about four households to exhaust coverage of the target population. The target population was defined to be all persons residing in each county at the time of the survey who were eligible to register. In the following discussion we refer to the above four frames collectively, for each county, as frame I and treat the frame for each county as a single frame. The second frame used in the two-frame sample design study was the county registration list, denoted frame II. Frame II was located in each of the eight county jurisdictions of interest and covered a portion of the target population. In this paper, we use the term multi-frame (two-frame) sample design to describe the design in which samples from both frames I and II are used in estimation of the registration and voting rates.

The initial goals in conducting the two-frame sample design were to test the feasibility of applying such a technique to the problem at hand and also to test if there were appreciable gains in efficiency of the design over other candidates. The potential for efficiency existed since in each county each person was required to register in order to vote. The registration list was used as a frame instead of a voting list because of its availability. The registration list also theoretically broadened the coverage of frame II over a voting list but with one serious drawback: the results of sampling from the registration list indicated moving rates of persons selected ranged from 20 to 50 percent. The voting list is believed to be more up-to-date.

The sampling procedure applied to frame II proceeded along the following lines. When possible, persons on the registration list were stratified using precinct information obtained from the official registrar in the county. That is, since the goal was to estimate characteristics by minority group, precincts with high concentration of minorities were grouped into strata and persons within sampled at differing rates among strata. At this time, the official (record checked) registration and voting information was recorded for the selected individual (sample person, SP) and every other person nearby on the list with the same address and surname. Then an interview was conducted at the given address and for all persons at the address, reported voting and registration information as well as minority status, etc. was recorded. Since the survey did not follow movers, any sampled person not found at the address (SPM) resulted in the exclusion of the entire interview at the household. Should the sample person be located at the address, the household was retained and each person interviewed was matched to the registration list to determine the weight to be applied to the household for estimation.

A subsample of households selected for the RAV survey was record checked against frame II and each person's official voting and registration status were recorded. Households were designated linkable if they contained at least one person who was found on the registration list. Otherwise, the household was termed nonlinkable. Clearly then each household selected from frame II was potentially linkable. The multi-frame feature of the designs evaluated below (models A and B) utilized this definition of linkable households (linkable to the registration list) in the estimation process.

Chronologically, the multiple frame sample design was implemented, data were collected, and estimates of the voting and registration rates were tabulated for both the reported and record checked data. Then the sample data were used to estimate various population parameters used in the model comparisons that follow. The analysis was based on estimating record checked voting and registration characteristics as the primary goal of each model.

In order to evaluate the cost efficiency of different survey designs that could have been used in the survey, five models were proposed. Models A and B used the same two-frame sample design but differed in their assumptions about the completeness of frame II. In Model B an up-to-date registration list was assumed in which all addresses were good while Model A retained the SPM. The estimation of the voting rate for models A and B is the ratio of the weighted total voting to the weighted total persons eligible to register to vote. These optimally chosen weights restricted to the range of 0 to 1. Specifically the numerator is the weighted total number of linkable households from both frames and the nonlinkable households from Frame I only. In models A and B record checking was assumed for both frames. Model C is a single-frame double sampling design which uses only the RAV sample and the RAV record-checked subsample. The ratio of the double sampling difference estimators is used to estimate minority voting rates. The reported data from the entire sample are used along with the reported and record-checked data from the subsample. Model D also uses only the RAV sample but no subsampling is assumed. The full sample is record-checked. In Model D the estimator of the voting rate is the ratio of the estimated total voting to the estimated total of persons eligible to register. The overall survey design for Model E is the same as for Model D but no record-checking is assumed. Examination of the data indicated that the square of the estimated bias from using only reported data exceeded the variance constraints. Model E was subsequently eliminated from succeeding comparisons.

Since a cost efficiency criterion was used to compare models, the optimum cost for each model was found for the same coefficients of variation. Optimization was done for both a single characteristic and for multiple characteristics. For a single characteristic the weights used in the Model A estimator are determined numerically by minimizing the cost function subject to a fixed coefficient of variation constraint. A typical cost function is a linear function of stratum sample sizes. The variance of the voting rate is inversely proportional to the stratum sample sizes with the coefficients being the stratum variances which are convex functions of the weights. Optimum stratum sample size is given by Neyman allocation. Survey costs with a given coefficient of variation were computed for several weight designations in the range of 0 to 1 and the optimum weights for a given coefficient of variation were selected to correspond to the minimum survey cost among the weights considered. Model D is a special case of Model A with zero weights applied to the linkage households from the registration list frame. The procedure to obtain optimum weights and minimum costs for Model B is the same as for Model A but adjustments were made to reflect an updated registration list. A Neyman allocation of the sample sizes was done for Model C which uses frame I only.

The resulting optimum sampling costs for each model and each single characteristic (specified minority voting or registration) were obtained for all eight counties. Since the optimum cost was quite different for each of the single characteristics, the overall model comparisons were conducted for multiple characteristics for each county. Causey's (1972) computer program was used for all model comparisons under multiple constraints. For Models A and B the weights derived from single characteristics were used.

On the basis of optimum survey costs for each model, Model C (the double sampling plan with a difference estimator) was the preferred model with Model D (1976 RAV sampling with all persons record-checked) or Model B (two-frame design with an updated registration list) second in preference. In two jurisdictions, Model B was the "best". Model B was

always better than Model A. This implies that if an updated registration list is available, survey costs can be reduced to a level where the two-frame approach may be able to compete with the single-frame approach. Given the registration and voting data and the sample designs in this study, the single-frame sampling approach proved to be better. Within the three single-frame models the double sampling approach with a difference estimator was better than the scheme requiring 100 % record-checking or the scheme requiring no record-checking in terms of cost per fixed coefficient of variation.

### 3.4 ON THE PROBLEM OF VARIANCE ESTIMATION FOR A DESEASONALIZED SERIES

K.M. WOLTER, N.J. MONSOUR, U.S. Bureau of the Census

#### ABSTRACT

The paper is concerned with the problem of variance estimation for a deseasonalized (or seasonally adjusted) time series. Statistical bureaus usually publish reliability statements and estimates of sampling variability associated with the original observations, but not for the deseasonalized observations. It is usually assumed either explicitly or implicitly that the variance associated with the adjusted series is equal to that of the original series. We discuss whether or not this assumption is valid, and propose methods for estimating the variances associated with the deseasonalized series.

The need for variance estimates for a deseasonalized series is not new, but rather has been recognized for at least 20 years. In the final report of the President's Committee to Appraise Employment and Unemployment Statistics (the Gordon Committee, 1962), it was suggested that standard errors for an adjusted labor force series are more important than those for the original series because of the increased reliance of policy makers on the deseasonalized series, and that research be undertaken on how to estimate and publish such standard errors. Unfortunately, this suggestion did not lead to any major advances or to practical methods for estimating and publishing the standard errors. As of the writing of this paper, most official publications of U.S. and Canadian statistical bureaus continue to provide the user with little information about the level of error in the deseasonalized statistics.

It is useful to distinguish between at least three different kinds of deseasonalized estimators, because the variance will depend on the kind of estimator. For ease of exposition, we relate this discussion to the Census X-11 algorithm (see Shiskin et al. (1967)), though similar remarks could be made about any seasonal adjustment methodology. In the U.S., it has been customary to project seasonal factors one year in advance using data through December of the previous year. As the observations are made for the new year, they are deseasonalized by using the projected seasonal factors. The resulting seasonally adjusted estimates are the first such estimates to be published for the new year, and are called the year-ahead deseasonalized estimates. A second kind of deseasonalized estimate is called the end-term estimate. This is prepared by using a seasonal factor that is computed from the observed series up to and including the observation being adjusted. The third kind of deseasonalized estimate, called the central-term estimate, is prepared by using a seasonal factor that is based on 145 observations (in the default version of the X-11) of the series, with 72 of the observations being on each side of the observation being adjusted. Clearly, this adjustment can only be made when the series is of length  $\geq 145$  and the observation being adjusted is at least 72 terms from the ends of the series. There are, in principle, 71 intermediate and asymmetric deseasonalized estimates that could be computed for a given time point  $t$ . In the U.S., it has been customary to run the X-11 program periodically computing the end-term adjustment, the central-term adjustment, and some of the intermediate adjustments. On occasion, a revised publication is issued, where the year-ahead deseasonalized estimates are replaced by later adjustments. Implicit in this discussion, of course, is a certain minimum length of the series. For series shorter than this minimum, X-11 operates in a different way, leading to further categories of deseasonalized estimates.

Although standard errors for deseasonalized estimators have apparently not been published nor the topic of widespread research, some limited research has been accomplished since the Gordon Committee's report (e.g. Kaitz (1974) and Monsour (1975)). This work showed that variances are roughly in the order Central-term adjusted  $\leq$  End-term adjusted  $\approx$  Original  $\leq$  Year-ahead adjusted.

In the paper, we discuss various approaches for the estimation of the variances of a seasonally adjusted series. We present two general models or concepts of the variance. One is the classical finite population model where the true series, i.e. the series that would result from a complete enumeration each month, is treated as fixed, and the variability of the deseasonalized values arises solely from survey errors. The second assumes a time series model for the true series, and the variability of the deseasonalized value arises from both survey errors and the irregular component in the series.

Some illustrative results are given based on various U.S. economic and demographic series for the first concept of variability. Two deseasonalization algorithms are considered. The first is the Census X-11 algorithm, which is based on ratio-to-moving-average methods. The second is based on a simple class of regression models. In the X-11 algorithm extensive use is made of Young's (1968) linear approximation to this algorithm. Graphical results are presented for both algorithms that confirm the earlier work about the rough order of the variances. Some practical variance estimation methods for deseasonalized series are suggested.

The second concept of variability is developed by assuming that the true series is generated by a linear regression model with time series errors. The total variance of the observed series is derived, and a strategy for variance estimation is suggested. This estimation strategy is applied to an inventory time series.

The main finding of this work is that the variance both for level and month-to-month change of a deseasonalized series is higher than that of the original series during the year-ahead months, but smaller during the central months of the series. The investigations in this paper are of a preliminary nature, and several extensions are already being planned. The methods extend to the optional filters, used in X-11 and to other deseasonalization algorithms, including Statistics Canada's X-11 ARIMA.

#### REFERENCES

- Kaitz, H.B. "Analysis of Current Economic Data on Employment and Unemployment". (1974). Unpublished manuscript prepared for the Committee on National Statistics, National Research Council, Washington, D.C.
- Monsour, N.J. "On the Analysis of the Standard Error of Seasonally Adjusted Series". (July 17, 1975). Unpublished memorandum to H. Nisselson, U.S. Bureau of the Census, Washington, D.C.
- President's Committee to Appraise Employment and Unemployment Statistics. Measuring Employment and Unemployment. (1962). U.S. Government Printing Office, Washington, D.C.
- Shiskin, J., et al. The X-11 Variant of the Census Method II Seasonal Adjustment Program. Technical Paper No. 15. (1967). U.S. Bureau of the Census, Washington, D.C.
- Young, A.H. "Linear Approximation to the Census and BLS Seasonal Adjustment Methods". JASA, 63, (1968), 445-471.

### 3.5 REDESIGN OF THE SURVEY OF RETAIL SALES (AUSTRALIA)

T. KARMEI, Australian Bureau of Statistics, Canberra, Australia

#### INTRODUCTION

The survey of retail establishments has been conducted by the Australian Bureau of Statistics since the early 1950's. The statistics of the retail industry provided by the survey are supplemented by a retail census which is conducted approximately every five years. The main aims of the census are to provide detailed cross-sectional data about the industry and to provide benchmarks for the survey. The latest retail census has been held in 1980, with respect to 1979/80.

The last major redesign of the survey was done in 1961 when a monthly survey was added to the quarterly. It has been felt necessary to evaluate the survey in terms of the information that it provides to users and the cost of the survey.

This paper outlines the sample design of the present survey, the desired changes to the present survey and the methodology which will be adopted to effect these changes.

#### DESCRIPTION OF THE PRESENT SURVEY DESIGN

The retail survey is made up of two surveys, a quarterly survey and a monthly survey. The quarterly survey is a stratified simple random sample of retail establishments. The stratification variables are state, area (metropolitan, ex-metropolitan) description (grocers, butchers, etc) and census retail sales. Retail sales for thirteen commodities are collected from all sampled establishments apart from those in the smallest size strata. Only total retail sales are collected for small establishments and commodity estimates are obtained from these units by applying commodity proportions, calculated using census data, to estimates of total sales at the selection stratum level.

Ratio estimation is used for all estimates. Within stratum ratio estimation is used for estimates of total sales and mainline commodity sales (the commodity which makes up the majority of sales in a description is termed the mainline commodity for that description) using, respectively, total sales and mainline commodity sales from the census as benchmarks. Across stratum ratio estimation is used for estimates of non-mainline commodity sales using non-mainline commodity sales from the census as the benchmark. For these estimates the ratio are formed at the state x description level.

The sample for the monthly survey, from which only total sales are collected, is a subsample of the quarterly sample. Two phase ratio estimation is used with the quarterly survey providing the first phase benchmark.

The retail survey, conceptually, should measure retail sales for all businesses in the retail sector. However, the survey only measures the sales by businesses actually on the survey framework. A new business provision must be added to the survey estimates in order to account for the sales by businesses not on the survey framework. The new business provision is based upon information from an annual postal listing and new businesses census. From the postal listing a list of new businesses is obtained by comparing a list of shops compiled by postmen with the businesses on the framework. All new businesses are sent a census form in order to obtain benchmark information so that the businesses can be included on the survey framework.

#### DESIRED CHANGES TO THE SURVEY

**Simplification.** The combination of two phase estimation, across and within stratum ratio estimation is felt to be excessively complicated. It has been found that simple designs are easier to maintain.

**Output of survey.** Users would like an improvement in the timeliness of quarterly results and an increase in detail for the monthly results.

**Resource saving.** There have been general pressures on the ABS in terms of resources. Most of the expense of the survey is in terms of processing survey returns.

**Response burden.** In the present survey units stay in the sample for five years. There are pressures on the ABS to minimize, especially for small businesses, the cost to businesses of filling in statistical returns.

**New business provision.** The present new business provision is unsatisfactory because it is not based upon a scientific sample survey. The problems are exacerbated by its size, and the fact that it is not possible to accurately measure the size of the errors. The poor quality of the provision has been manifested by the need to have substantial revisions to the provision.

#### PROPOSED METHODOLOGY FOR THE REDESIGNED SURVEY

There are two major changes to the sample design which have been proposed. The first is the use of synthetic estimation for all businesses, apart from very large businesses, by combining census commodity information with total sales figures from the survey. The second is a new approach to estimating the new business provision by using a sample of census collectors districts to obtain a probability sample of new businesses. Other less important changes have been proposed, including an increased use of ADP editing and the use of a collocated selection scheme to give some control over response burden.

**Synthetic Estimation.** The idea behind using synthetic estimation is to reduce the amount of data collected in the survey by combining the survey total sales data with census commodity sales data to produce estimates of commodity sales for the survey reference period.

A detailed investigation showed that the mean square errors of the synthetic estimates would not be much different from the variances of the survey estimates based on actual commodity data. This result is influenced by the decision to collect commodity data from all businesses in the completely enumerated sector.

The immediate implication of using synthetic estimation to estimate commodity sales is that it cuts down the amount of data that has to be collected and processed. This enables simplification of the survey design and an increase in the detail published. The proposed changes are :

- the same sample design and sample for the monthly and quarterly surveys
- estimation of total sales using ratio estimation with census total sales as the benchmark
- publication of total sales, by state and description, every month
- quarterly publication synthetic estimates of commodity sales

New business Provision. It is proposed (but not yet decided) to replace the annual postal listing and new business census with a quarterly area sample and new business survey. The aim of the area sample is to provide an up to date sample of businesses not on the retail survey framework. These new business would be the sample for the new business survey which would provide a direct estimate of sales by businesses not on the retail survey framework.

The area sample would be a stratified sample of census collectors districts. The stratification variables would be geographic region and number of shops in the CD. The sampled CD's would be listed by field agents to obtain a list of new businesses which would be sent retail survey forms. Usual updating sources of the Integrated Register (from which the survey framework is derived) would also provide some new businesses which would also be included in the new business survey.

The sample of CD's would be selected independently once a year so that the new business found by the field agents could be updated to the survey framework for the annual sample revision.

Response burden. Because estimates of movement are important it is not desirable to rotate the sample. However in order to ease the burden of small businesses it has been proposed that the businesses in the smallest size strata should be rotated by 25 % every quarter. A collocated sampling scheme will be used to effect the rotation.

CONCLUSION

It is intended to introduce major changes to the retail survey. These changes are intended to simplify the survey, to save resources and at the same time improve the range of data produced by the survey. Also the quality of the new business provision will be greatly improved. Development work is proceeding at the moment and it is intended to introduce the redesign at the beginning of 1982.

3.6 RECTIFICATION ANNOUNCEMENT

" Bias and non-response rate reduction using a randomized response model among the nonrespondents", C.N. BOUEA (Cuba)

A revised version of this paper published in the Survey Statistician n° 5 will be inserted in the next issue of the Survey Statistician.

4. NEWS FROM THE ASSOCIATION

4.1 OBITUARY

Professor Jerzy Neyman died this summer at the age of 87 years.

It is difficult, or perhaps impossible, to give a brief assessment of the role that Neyman played in the development of statistical theory and methods for more than 50 years. Together with Egon Pearson, he laid the foundations of one theory of statistical inference, properly known as the "Neyman-Pearson theory".

In recent years, there has been considerable discussion of the gap between statistical theory and applications. In our endeavors to close that gap, we can learn much from Neyman's applied work. Neyman applied his theory to a great variety of problems in various fields : agriculture, astronomy, environmental protection, medicine, to mention but four. It is of special interest to note that his applied work played an instrumental role in his research : he identified problems, which he then tackled with great skill and ingenuity, reflecting his thorough training in mathematics.

To the members of IASS, Neyman will be especially remembered for his pioneering contributions to survey sampling theory and methods. He is the founder of what may be called the probability survey sampling school. We all admire his work in the realm of stratified sampling, first published in Polish (in 1933). His 1934 paper in the Journal of the Royal Statistical Society, "On the two different aspects of the representative method : the method of stratified sampling and the method of purposive selection" was a breakthrough for modern ideas. More important, in my view, than his discussion of "minimum variance allocation", also referred to as "Neyman allocation", was his successful effort to place the theory of survey sampling in the framework of general statistical inference theory.

As said in the introduction, Neyman played a decisive role in the formation of statistical theory and methods over more than half a century. His success in his endeavors was, of course, primarily due to his scholarly characteristics. But an important factor was his great personal characteristics : he was generous with his help and his time ; not least important was it that he always tried to be fair in discussions about his contributions and those of others ; he could admire a theoretical contribution not only from followers, but also from scholars with different views about the choice of groundwork for statistical theory and methods.

Jerzy Neyman is no longer among us. But what he contributed to our discipline will be with us for ever.

T. DALENIUS  
President, IASS

4.2 BUENOS AIRES SESSION, ARGENTINA, NOVEMBER 30 - DECEMBER 11, 1981

4.2.1. Meetings organized jointly by IASS and ISI

Title	Organizer	Invited papers
Conceptual and theoretical framework for survey sampling (n°12)	N.P. Purcell (Australia)	12.1 - The theory and practice of survey sampling (C.O'Muircheartaigh, T.Wong)
		12.2 - Conceptual frameworks in survey sampling with applications to the non-response problem and the small area estimation problem (K.E. Sarndal)
		12.3 - Models in the practice of survey sampling (G. Kalton)

Discussant : L. Kish

Title	Organizer	Invited papers
Analysis of data based on complex survey designs (n° 13)	G. Nathan (Israel)	13.1 The analytical use of unequal probability samples : a case study (K. Brewer) 13.2 Analysis of Canada health survey data (J.N.K. Rao, M. A. Hidiroglou) 13.3 Estimating victimization prevalence in a rotating panel survey (W.F. Eddy, S.S. Fienberg, D. Griffin)
Discussants : C. O'Muircheartaigh C. Sarndal		
Evaluation of the quality of statistical data (n° 14)	B.A. Bailer (USA)	14.1 Evaluation of some aspects of the Israeli family expenditure survey (M. Kantorowitz) 14.2 1980 US Census population coverage (C.D. Jones) 14.3 Plans and considerations in evaluating the 1982 census of population and housing in Morocco (A.L. El Ghazali, H. Hogan)
Discussants : G. Bravo J.R. Coker		
Household sample surveys in developing countries with emphasis on non-sampling errors (n° 15)	L. Diop (Senegal)	15.1 The technical problems of household surveys in selected countries (J. Nieto de Pascual) 15.2 Technical problems of coverage and content error in household surveys (V. Verma)
Discussants : L. Kish, W.L. Booker, S.S. Zarkovich, K. Krotki, M.N. Murthy		

#### 4.2.2. Summary booklet

IASS intends to publish as usual, a summary booklet including summaries of the invited papers presented at the joint IASS-ISI meetings. In that purpose, authors of the concerned papers are kindly requested to mail at earliest two copies of a two-page summary, in English or French, to the IASS Secretariat, to Mr Damiani, c/o INSEE, 18 boulevard A. Pinard, 75675 Paris Cedex 14, France.

The booklet will give too the list of the contributed papers presented during the meetings.

#### 4.3 IASS PROGRAM COMMITTEE, 1983 MADRID SESSION

The composition of the IASS Program Committee for the 1983 session in Madrid is as follows :

##### Chairman

C.O'Muircheartaigh, London School of Economics, Houghton street, London WC2A 2AE, U.K.  
N.B. : from July 1981 to July 1982, Mr O'Muircheartaigh will be located at the Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106, USA.

##### Committee members

- B.A. Bailer (Bureau of the Census, USA)
- B. Kiregyera (University of Makerere, Uganda)
- A.L. Finkner (Research Triangle Institute, USA)
- G. Nathan (Hebrew University of Jerusalem, Israel)
- N. Purcell (Australian Bureau of Statistics, Australia)
- J.L. Sanchez-Crespo (Instituto Nacional de Estadística, Spain)
- D. Singh (Indian Agricultural Statistics Research Institute, India)
- G. Théodore (INSEE, France)
- V.K. Verma (United Nations Statistical Office, USA)

Any member wishing to suggest subject areas, topics, proposals for meetings and names of potential organizers should write to the Chairman or to any member of the Committee as soon as possible. The program will be finalized during the 1981 session in Buenos Aires.

#### 4.4 NEW IASS CORRESPONDENTS

IASS is grateful to the two people who accepted to serve as correspondents for the account of the Association :

- in Brazil : Mr de Oliveira Bussab, Dep. estadística, IME-USP, Caixa Postal 20570, 05508 Sao Paulo, Brazil
- in USA , to replace Ms B. Boyes who died lately : Mr C. Jacobs, Chief of Division of Consumer, Statistical Methods and Evaluation Research, GAO Building, 441 G street N.W., Rm 2122, Washington D.C. 20212, USA

## 5. BIBLIOGRAPHY

- W.G. HUNTER - "The practice of Statistics : the real world is an idea whose time has come" - The American Statistician, May 1981, vol. 35, N°2
- J.C. KOOP - "The problem of normality in statistical inference for sample surveys" - Research Triangle Institute, proceedings of the American Statistical Association, 1980, Section on Survey Research Methods

J. FRANKEL, L.M. SHARP - "Measurement of respondent burden" - Summary of study design and early findings - Statistical Reporter, January 1981, N° 81 - 4 - US Department of Commerce

SURVEY METHODOLOGY - TECHNIQUES D'ENQUETES - Vol. 6, N°2, December 1980 - Statistique Canada

Special issue :

- Data, statistics, information - Some issues of the Canadian social statistics scene, I.P. FELLEGI
- Sampling with unequal probabilities and without replacement. A rejective method, G.H. CHOUDHRY, M.P. SINGH
- Test of multiple frame sampling techniques for agricultural surveys : New Brunswick, 1978, B. ARMSTRONG
- Canadian victimization surveys : a report on pretests in Edmonton and Hamilton; G. CATLIN, S. MURRAY
- A personal view of hot deck imputation procedures, I.G. SANDE

SURVEY METHODS NEWSLETTER - Spring 1981 - SCPR (Social and Community Planning Research) - London

Contents :

- The report on the March 1981 SSRC Survey Methods Seminar on "Sampling Minority Populations"
- Programme for the November 1981 SSRC Survey Methods Seminar
- Outline of SCPR Survey Methods Centre's programme of research on interviewing methods

NATIONAL MIGRATION SURVEYS - SURVEY MANUALS - Economic and Social Commission for Asia and the Pacific, U.N. New York 1980

- Survey organization and monitoring
- The core questionnaire
- Manual for interviewers
- Manual for supervisors
- Manual for interviewer trainers
- Sample design manual
- Manual for office editors and coders

SURVEY RESEARCH - Vol. 12, N° 13 - Summer 1980 - Survey Research Laboratory, University of Illinois, USA

STATISTICAL POLICY WORKING PAPER 6 - REPORT ON STATISTICAL USES OF ADMINISTRATIVE RECORDS - US Department of Commerce, Office of Federal Statistical Policy and Standards, December 1980

POPULATION BULLETIN OF THE UNITED NATIONS ECONOMIC COMMISSION FOR WESTERN ASIA, N° 17, December 1979

POPULATION BULLETIN OF ECWA - N° 18, June 1980 - Economic Commission for Western Asia, UN . Beirut, Lebanon

DEMOGRAPHIE AFRICAINE, N° 34, novembre-décembre 1980 - Institut de formation et de recherche démographiques (I.Fo.R.D.)

ASIAN AND PACIFIC CENSUS FORUM - Vol. 7, N°3, February 1981 - East-West Population Institute, East-West Center, Honolulu, Hawaii

MEMBERSHIP APPLICATION FORM

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

NAME : Mr.  First name (s) .....  
Mrs.   
Miss

Date of Birth : ..... Nationality : .....  
Profession : ..... (if retired, please indicate)

Title (professor, doctor) : .....

Business address :

Name of office : .....  
Street : ..... N° : .....  
City : ..... Postal code : .....  
Country : .....  
Telephone n° : ..... Extension : .....

Home address :

Street : ..... N° : .....  
City : ..... Postal code : .....  
Country : .....  
Telephone N° : .....

IASS correspondance should be sent to :

- Business address
- Home address

Your mother tongue is : .....(Please specify)

The official language of your IASS correspondance will be :

- English
- French

I am aware that the membership dues have been fixed at 10.00 U.S. dollars (or the equivalent in other currencies) and that, exceptionnally for nationals of developing countries, these dues have been reduced to 5.00 U.S. dollars.

Modes of payment

- Cheque payable to the International Association of Survey Statisticians \*
- UNESCO coupons (for countries with inconvertible currencies) \*
- Bank account of the I.A.S.S. N° 4586-32 with the Banque Nationale de Paris - Agence Didot, 67 boulevard Brune, 75014 PARIS, France
- Other

\* Sent directly to : Mr. Paul DAMIANI, IASS Secretary  
c/o INSEE  
18, boulevard Pinard  
75675 PARIS CEDEX 14 - France

I have been recommended by :

Signature :

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, organizations, institutions, governments and the general public in different countries of the world.

OFFICERS AND MEMBERS OF THE COUNCIL

President (1980-81)  
President elect (1980-81)  
Executive Director  
Vice-presidents (1980-81)

Scientific Secretary (1980-81)  
Treasurer (1980-81)  
Director of the ISI Permanent Office  
Members (1980-81)

(1980-83)

T. DALENIUS (Sweden)  
S.S. ZARKOVICH (Yugoslavia)  
G. THEODORE (France)  
J. DESABIE (France)  
L.E. ROWEBOTTOM (Canada)  
A.L. FINKNER (USA)  
G. THEODORE (France)  
E. LUNENBERG (Netherlands)  
W.E. DEMING (USA)  
C. HAYASHI (Japan)  
D. HOLT (United Kingdom)  
K.J. KROTKI (Canada)  
G. NATHAN (Israël)  
A.S. NIELSEN (Denmark)  
E.K. FOREMAN (Australia)  
K.T. de GRAFT-JOHNSON (Ghana)  
G. KALTON (United Kingdom)  
M.N. MURTHY (India)  
H. NISSELSOON (USA)  
J.N.K. RAO (Canada)

SECRETARIAT

Mr. Paul DAMIANI, I.A.S.S., 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.

It contains practical information in sampling surveys : general information (proceedings of session and surveys, surveys in progress), papers (survey results, sampling technology), terminology, question/answer, education, news of the Association, bibliography.

Editor : J. Desabie, Inspecteur Général de l'INSEE, INSEE, 18 bd Adolphe 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).

