INTERNATIONAL STATISTICAL INSTITUTE

INTERNATIONAL ASSOCIATION OF SURVEY STATISTICIANS

THE SURVEY STATISTICIAN

n°19

C O N T E N T S

*****

JULY 1988

I - A MESSAGE FROM THE PRESIDENT

It is indeed an honour for me to have been elected president of the I.A.S.S. It will be particularly difficult to emulate the achievements of my predecessors, all of whom have given sterling service to the Association. With the help of the Executive Director, M. Charoy, and the Secretariat, and with the participation of the members of the Council, I hope that the next two years can be a continuing period of development for I.A.S.S.

I would like to take this opportunity to redress an oversight on our part. We have not so far given due recognition to the Swedish for Research Cooperation with Developing Countries (SAREC) and Statistics Sweden, who printed and distributed free of charge Elements of Survey Sampling by Tore Dalenius to members of I.A.S.S. living in developing countries. We are grateful to them for this contribution, and are pleased to say that some supplies of the book are still available, and a copy will be distributed to new I.A.S.S. members in developing countries as long as stocks last.

It gives me great pleasure to express our appreciation to the Secretariat and to I.N.S.E.E. for producing so promptly the booklet containing the papers on survey statistics given at the Tokyo meetings. This volume should now have been received by all members of the I.A.S.S. who are not members of the I.S.I. and constitutes a very important service to the membership. The booklet has been issued with the agreement of the I.S.I. Permanent Office and the Japanese Organizing Committee. We thank them both.

Following the successful workshop for survey statisticians from developing countries organized by Dr. D. Holt at the Tokyo meetings, we intend to organize a similar session immediately preceding the I.S.I. meetings in Paris in 1989. The workshop will be coordinated by Dr. Barbara Bailar, the I.A.S.S. President-Elect. A notice about the workshop appears in this issue of the Survey Statistician.

Dr. Graham Kalton, Vice President, has undertaken a great deal of work in updating and improving the materials available for local representatives. A new package of materials is being prepared and will be sent shortly to all local representatives. We are grateful to Dr. Kalton for his efforts in this matter.

M. Alain Lery is the Chairman of the I.A.S.S. Nominations Committee. Any suggestions about members who should be nominated for positions either as officers or as members of the Council should be sent to him direct at I.N.S.E.E. (18, boulevard Adolphe Pinard, 75675 PARIS CEDEX 14).

Colm O'Muircheartaigh

President of the I.A.S.S.

II - NEWS FROM THE ASSOCIATION

2.1. Brief news from the Council

Dr. Kirk M. Wolter of the U.S. Bureau of the Census will be Programme Chairman for the 1981 meetings in Cairo. As he has already begun work on the preparation of the programme, suggestions about topics, speakers, discussants and organizers for the 1981 sessions should be sent directly to him, or to the Secretariat who will forward any materials to him.

A meeting of the Executive will take place in Paris later this year. Any members who have issues that they would like to have raised at the Executive meeting should communicate either with the President Colm O'Muircheartaigh or directly with M. Charoy at the Secretariat. The development of the Association depends greatly on the participation of the members, and we would therefore welcome any suggestions, criticisms or contributions from members on any issue related to the work of the Association.

2.2. 1989 Paris Session

List of invited papers meetings of the 47th I.S.I. Session (1989) supervised by I.A.S.S. (with the names of the organizers approved by the I.S.I. Programme Co-ordinating Committee).

TOPIC ORGANIZER

N. 10"Data collection and estimations for informal Economies" J. CHARMES (France)

N.12 "Methodology of Household Budget Survey"V.K. VERMA (U.K.)

N.15 "Methodology of Agricultural Survey"E.O. BOATENG (Ghana)

N.16 "Reduction of Memory Errors in retrospective Surveys" D. SIKKEL (Netherlands)

N.17 "Estimation and Analysis of Panel Survey Data"L. FABBRIS (Italy)

N.18 "Quality improvement of Survey Questionnaires"T. WALCZAK (Poland)

N.19 "Survey Evaluation with special reference to non-sampling errors"D. TREWIN (Australia)

- The meeting N.10 should be held with the cooperation of I.A.O.S.

- The I.A.S.S. should be cooperate in the following meetings :

N. 7 "Methods for the Analysis of Complex Socio-Economic Data", under the responsability of I.S.I.

N.14 "Automation of Survey Data : Collection and processing", under the responsibility of the I.A.S.C.

List of names and addresses of organizers of I.A.S.S. sponsored meetings.

J. CHARMES Institut National de la Statistique et des Etudes Economiques
Departement de la Cooperation
18, Boulevard Adolphe Pinard
75675 PARIS CEDEX 14
TELEX N. 204924 F INSEE

V.K. VERMA 105, Park Road
TEDDINGTON (Middlesex) - TWLL OAW (United Kingdom)

E.O. BOATENG Government Statistician Central Bureau of Statistics
P.O. Box 1098
ACCRA (Ghana)
CABLE : GHANASTATS

D. SIKKEL Schout van Eijklaan 98
2262 XV Leidschendam
(The Netherlands)
TELEX N. 26383 RESIN NL

L. FABBRIS Dipartimento di Scienze Statistiche
Università di Padova
Via S. Francesco, 33
35121 PADOVA
TELEX N. 430176 UNPADUI

T. WALCZAK Vice-Pres of the Statistical Office of Poland
Al Niepodleglosci 208
VARSAVIA 58
TELEX N. 814581C GUS PL

D. TREWIN Australian Bureau of Statistics
P.O. Box 10
BELCONNEN A.C.T. 2616 (Australia)
TELEX N. ABOST AA61872

2.3. Workshop for survey statisticians-Paris 1989

The International Association of Survey Statisticians (I.A.S.S.) will sponsor a workshop for survey statisticians from developing countries immediately preceding the meeting of the International Statistical Institute (I.S.I.) in Paris, France on August 27 - 29, 1989.

The expenses for some participants to attend the workshop and the I.S.I. will be paid for by the I.A.S.S., through funding provided by companies and agencies worldwide. Persons interested in attending should prepare a two or three page summary describing an aspect of their work. The summary should include one or more specific issues or problems. Some examples of broad topics are : agricultural surveys, household surveys and policy development. Examples of issues are : choice of frame for sampling, sampling unit, estimation and weighting. The summary should be as specific as possible so that resource people with experience in the topic can be recruited to help at the workshop. Participants will be selected on the basis of this summary.

The language of the workshop will be English. No translation facilities will be available. The workshop will be coordinated by Dr. Barbara Bailar, President-Elect of I.A.S.S.

Statisticians who would like to be considered for inclusion should write as soon as possible to Dr. Bailar at the address given below. Decisions on applications will be made in early 1989 and applicants should respond without delay to this announcement.

Applications should include :

(I) A brief outline (500 words) of a suitable topic for the Workshop.

(II) As much information as possible on travel costs from the applicants home to Paris.

(III) A statement describing any financial support already available to the applicant.

(IV) A brief description of education and work experience including current position.

(V) Adress, telephone and telex numbers to enable fast communication.

Applications should be sent to :

Dr. Barbara A. Bailar
Executive Director
American Statistical Association
1429 Duke Street
Alexander, VA 22314-3402

2.4. "Survey Methodology"

Members of I.A.S.S. who subscribed to the Journal for 1987 will receive two issues of Volume 13. The June 87 (no. 1) issue was mailed to members in December 1987, and the December 87 (no. 2) issue will be mailed about April 1988. Statistics Canada is trying to make the dates of publication conform more closely with the date on the issues. Delays may also be caused if members' addresses that we have are not up-to-date. Please notify the Production Manager, Survey Methodologyn Statistics Canada, Ottawa, Ontario, Canada K1A OT6 when you change your address.

The 1988 issues will include special sections devoted to the estimation of Census coverage error. In order to qualify for the special reduced rate of US $ 10 (US $ 5 for developing countries) for subscriptions to Survey Methodology, members must order through I.A.S.S., and subscriptions must be renewed through I.A.S.S. each year.

Survey Methodology

A Journal of Statistical Development and Applications in Survey

Management Board : G.J. Brackstone (Chairman), B.N. Chinnappa, G.J.C. Hole, C. Patrick, R. Platek, D. Roy, M.P. Singh, F. Mayda (Production Manager).

Editorial Board : Editor - M.P. Singh ;
Associate Editors - K.G. Basavarahappa, D.R. Bellhouse, L. Biggeri, D. Binder, E.B. Dagum, W.A. Fuller, J.F. Gentleman, D. Holt, G. Kalton, M.N. Murthy, W.M. Podehl, J.N.K. Rao, I. Sande, C.E. Särndal, F.J. Scheuren, V. Tremblay, K.M. Wolter ;
Assistant Editors-J.Armstrong, J. Gambino, J.L. Tambay.

Editorial Policy : SURVEY METHODOLOGY publishes articles dealing with various aspects of statistical development relevant to a statistical agency, such as design issues in the context of practical constraints, use of different data sources and collection techniques, total survey error, survey evaluation, research in survey methodology, time series analysis, seasonal adjustment, demographic studies, data integration, estimation and data analysis methods, and general survey systems development. Emphasis is placed on the development of specific methodologies as applied to data collection or the data themselves.

All articles appearing in SURVEY METHODOLOGY are published in both English and French.

Invitation to Authors : are invited to submit manuscripts in either English or French to SURVEY METHODOLOGY. All manuscripts are refereed. Two copies should be sent to : Editor, Survey Methodology, Methodology Branch, Statistics Canada, Ottawa, Ontario, Canada, K1A OT6.

Subscription Information : SURVEY METHODOLOGY is published twice a year. Annual subscription (Statistics Canada catalogue no. 12.001) is $20.00 in Canada, $23.00 elsewhere (payment to be made in Canadian funds or equivalent).

Special Price for I.A.S.S. Members : For members of the International Association of Survey Statisticians, the American Statistical Association and the Statistical Society of Canada, SURVEY METHODOLOGY is available at the reduced price of US $10.00 ($14.00 Can.). To receive this reduced price, please subscribe through your Association. I.A.S.S. members should send subscriptions to :

I.A.S.S. Secretariat,
c/o INSEE, 18, boulevard Pinard,
75675 Paris Cedex 14,
France

2.5. Local Representatives

The new Local Representative for Australia is :

David Steel

Australian Bureau of Statistics

P.O. Bos 10

Belconnen 2616 - AUSTRALIA

2.6. Competition for young statisticians from developing countries

The International Statistical Institute (I.S.I.) announces the Fourth Competition among young statisticians from developing countries who are invited to submit a paper on any topic within the broad field of statistics, for possible presentation at the 47th Session of I.S.I. to be held in Paris, France, in 1989.

Participation in the competition is open to nationals of developing countries who are living in a developing country, who will not be older than 32 years of age in the year during which the Session is to be held.

Papers submitted must be unpublished, original works which may include university theses.

The papers submitted will be examined by an international Jury of distinguished statisticians who are to select the three best papers presented in the competition. Their decision will be final.

The authors of the winning papers will be invited to present personnally their papers at the Session of I.S.I. concerned with all expenses paid (i.e. round trip airline ticket from his/her place of residence to Paris plus a lump sum to cover living expenses).

Manuscripts for the Competition should be submitted in time to reach the I.S.I. not later than November 1, 1988.

The rules governing the preparation of papers, application forms and full details are available on request from the I.S.I. Permanent Offices to which interested individuals should write for further information. The address is as follows :

The Director

Permanent Office

International Statistical Institute

428 Prinses Beatrixlaan

2270 AZ Voorburg

The Netherlands

III - ANNOUNCEMENTS

Fifth annual research conference of the U.S. Bureau of the Census (ARC V) (March 1989 - Washington D.C.)

The Bureau of the Census is planning its Fifth Annual Research Conference, to be held in March of 1989 in the Washington, DC area. The conference will consist primarily of contributed papers, most of which receive formal discussion at the conference. The conference will feature papers on topics related to a broad range of Census Bureau research interests. Papers may address methodology, empirical studies, or relevant issues. A conference proceedings volume containing all paper and discussions will be published. Papers must be original and not previously published or disseminated. Presenters will be reimbursed for transportation and per diem expenses and will receive a fee for manuscript preparation (expected range : $250--$450).

Topic areas include :

* sample design for household surveys and basic designs for labor force surveys
* new techniques in questionnaire design, including cognitive research and dependent interviewing
* statistical analysis of complex survey data
* managing the changing survey workforce, including recruiting, training, and supervision
* modeling survey costs, including model development, validation, and estimation methods
* mapping and geographic systems, including spatial analysis and text placement algorithms
* generalized computer systems for edit and imputation of survey data
* longitudinal analysis of establishment and household survey data
* quality and coverage of foreign trade statistics, including data reconciliation agreements, and scope of sector coverage
* maintaining establishment and business registers, including methods, technology, and systems
* disclosure avoidance techniques
* managing large statistical data bases containing past as well as current data
* record linkage methods, software, and applications
* contributions of organizational research to statistical organizations
* non-sampling errors

To have a paper considered for presentation, send a 500-word abstract by June 1, 1988 to :

David F. Findley
Conference Chair
Statistical Research Division
Bureau of the Census
Washington, DC 20233

To obtain registration information or to be included on the mailing list, contact.

Maxime Anderson-Brown
Conférence Coordinator
Office of the Director
Bureau of the Census
Washington, DC 20233

Please note, plans for ARC V are dependent upon final approval and funding which are still pending.

IV - COUNTRY REPORT

Smallholding crop surveys in Papua New Guinea : how good was the sampling design ?

1. Background

Papua New Guinea, with a population of 3 million in 1980, lies just south of Equator and is situated between South East Asia and Australia. It has 19 administrative provinces - with Port Moresby as the National capital. During the period 1979-84, a series of surveys were conducted by the Rural Statistics Section of the Department of Primary Industry in 11 provinces of the country in order to provide baseline information on smallholding agricultural activities. This was the first major attempt in recent years to obtain detailed information on crops at provincial levels ; a previous survey of a similar nature, conducted during 1975/76, provided rather limited information at provincial levels. Some of the results have been described in a previous paper by Koley.

2. Sample design

A stratified two-stage sampling scheme was used in all provinces (with minor variations in three). Districts formed the geographical stratification in case of 8 provinces. In case of Milne Bay province, the stratification was extended at Census Division levels. During the last two provincial surveys, i.e., North Solomons and Southern Highlands provinces, however, another level of stratification was considered on the basis of agricultural activity (i.e., participation in cash crop production). Within each stratum thus defined, primary units (villages or census units) were selected with a probability proportional to estimated size (population in most cases) or "p.p.e.s." with replacement. From within each selected primary unit, a requisite number of second stage units or sub-units (household) was selected randomly without

replacement. Information collected were on number of households, gardens, treeswith a breakdown according to maturity as well as numbers affected by pests and diseases with respect to export crops. Also some information on food crops was collected at household levels.

The sample size was determined mainly on administrative and financial considerations rather than sampling efficiency. A wide range of variation in estimates of sampling errors was, not unexpectedly, observed. Table 1 provides a comparison of co-efficients of variation between provinces in respect of a few major characteristics (e.g., numbers of coffee, cocoa and coconut trees).

3. Sample estimates and variances

As has been mentioned earlier, the sample scheme adopted was a stratified two-stage one. The formulae for the estimates and their variances have been drawn from Cochran 2 and are briefly presented in the following sections. For the sake of convenience, sampling within a stratum is considered first.

Let n = no. of sample primary units ;

N = no. of population primary units ;

m_i = no. of sample sub-units within ith selected primary unit ($i = 1, 2, \ldots\ldots n$) ;

M_i = no. of corresponding population sub-units ($i = 1, 2, \ldots\ldots n ; n+1, n+2, \ldots\ldots N$) ;

M_o = Total no. of population sub-units = $\sum M_i$;

y_{ij} = value of characteristic for jth sub-unit within ith primary unit ($i = 1, 2, \ldots\ldots n ; j = 1, 2, \ldots\ldots m_i$) ;

$$\bar{y}_i = \frac{1}{m_i} \sum_{j=1}^{m_i} Y_{ij} \; ; \; Y_i = \sum_{j=1}^{M_i} Y_{ij} \; ; \; Y = \sum_{i=1}^{N} Y_i$$

p_i = probability or relative size assigned to the ith unit ($\sum p_i = 1$)

Now, the unbiased estimate of the population total Y is :

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^{n} \frac{M_i \bar{y}_i}{p_i} = \frac{1}{n} \sum_{i=1}^{n} k_i = \bar{k} \qquad \text{(say)}$$

where $k_i = \dfrac{M_i \bar{y}_i}{p_i}$ ..... (3.1)

and its variances is :

$$V(\hat{Y}) = \frac{1}{n} \sum_{i=1}^{N} p_i \left\{ \frac{Y_i}{p_i} - Y \right\}^2 + \frac{1}{n} \sum_{i=1}^{N} \frac{M_i(1-f_{2i}) S^2_{2i}}{m_i p_i}$$

where $f_{2i} = \frac{m_i}{M_i}$

and $S^2_{2i} = \frac{1}{M_i - 1} \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_i)^2$

where $\bar{Y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} Y_{ij} = \frac{Y_i}{M_i}$ .....(3.2)

An unbiased sample estimator of V (Y) is :

$$v(\hat{Y}) = \frac{\sum_{i=1}^{n} (k_i - k)^2}{n(n-1)} \qquad .....(3.3)$$

The provincial estimates and teir variances can now be obtained by simply adding strata estimates.

Some interesting situation occurred when a primary unit was selected more than once. Though classically it is supposed that, on each selection, the whole subsample should be replaced and a new independent drawing of m_i sub-units be made without replacement from the complete unit, in practice, however, this posed a major administrative problem. The alternative used was, therefore, to draw a single sub sample of size m no matter how many times the ith unit is selected. The estimate, $k_i = \frac{M_i \bar{Y}_i}{p_i}$ from this unit received a weight t_i (no. of times the ith unit is selected).

The effect of this modification would be to increase $V(\hat{Y})$ by

$$\frac{n-1}{n} \sum_{i=1}^{N} \frac{M_i^2 (1 - f_{2i}) S^2_{2i}}{m_i} \qquad ..... (3.4)$$

However, for the same cost, the difference in precision between these methods are seldom likely to be substantial according to Cochran.

4. Validity of sampling design

The rationale for p.p.e.s. sampling with replacement (at the first stage) was based on three major assumptions :

a) if the regression of the variable to be measured (y) on the measure of size (x) is found to be a straight line passing through the origin, probability proportional to size (or estimated size) will be very efficient according to Murthy[4].

b) in a multi-stage sampling scheme, selection of primary units with probabilities proportional to a measure of size (x_i) is at its most effective, relative to selection with equal probabilities when the ratios of primary unit totals to the sizes, (ie. y_i / x_i) are uncorrelated with the sizes (x_i) for the principal items for estimation according to Cochran[5].

c) the gains in precision from sampling without replacement will be relatively smaller in a two-stage sampling. Furthermore, the formulae for the variances of the estimates in a multi-stage sampling with replacement where the primary units are selected with p.p.e.s. are decidedly simpler when compared with those in sampling without replacement or simple random sampling.

5. Tests in Southern Highlands province

A few tests were carried out in the two provincial crop surveys, i.e., Southern and North Solomons to check whether the first two assumptions hold good. As has previously been mentioned, in case of Southern Highlands province, a further stratification (apart from administrative) was made of all villages according to the proportion of households growing coffee in the village. The villages in which 25 % or more of households were coffee-growers (according to 1980 Census) were regarded as "Coffee" villages whereas the rest were considered as "Non-Coffee" ones. Out of a total of 121 villages surveyed, 96 were found to be "Coffee" villages and 25 were "Non-Coffee".

For the purpose of the test, the major variable considered (y) was the total numbers of trees and the two hypothesis as mentioned in (a) and (b) were considered for testing.

The hypothesis (a) has two components, i.e., in a linear equation : $Y = \alpha + \beta x$, it is to be tested whether i) the regression is linear, i.e., $\beta = 0$, and ii) the line of regression passes through the origin i.e., $\alpha = 0$.

In both cases, however, the null hypothesis should be tested against the alternative.

First to test the hypothesis i) $H_o : (\beta = 0)$, the corresponding t-statistic is :

$$t_b = \frac{b}{S_{y.x} / \sqrt{\sum (x_i - \bar{x})^2}} \quad \text{with } n - 2 \text{ d.f}$$

where $S^2_{y.x} = \dfrac{\sum (y_i - a - bx_i)^2}{n - 2}$ ..... (4.1)

and the t-statistic for testing the hypothesis ii) Ho : ($\alpha = 0$) is :

$$t_a = \frac{a}{s_{y.x} \Big/ \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum (x_i - \bar{x})^2}}} \quad \text{with } n - 2 \text{ d.f} \qquad \ldots.(4.2)$$

Where a and b are least-square estimates of α and β respectively in the linear regression equation. The above tests have been derived from Snedecor & Cochran[6].

The results of the tests are summarised in table :

Table 1 : Tests for linearity of regression in Southern Highland Province

Stratum	d.f	a			b		
		Value	t	Status	Value	t	Status
Coffee	94	668	0.045	I	422	3.539	HS
Non-Coffee	23	17,812	1,493	I	89	0.816	I
TOTAL	119	2,462	0.201	I	346	3.454	HS

(Note : I = Insignificant at 5 % level ; HS = Significant at 1 % level).

From the above table, it is apparent that the value of b is highly significant both for total as well as "coffee" stratum though it is insignificant for "non-coffee" which means there is a linearity of regression of y on x in coffee growing areas, but this is lacking in "non-coffee" areas. This result is consistent as x would be expected to have relatively less association with y in "non-coffee" areas. It is further observed that the value of "a" on the other hand is consistently insignificant, showing thereby the intercept of regression line, if there is one, is not significantly different from zero.

Next, to test the hypothesis (b), correlation (r) between x (no. of households) ans y/x (no. of trees per household) was considered and the results are summarised as follows :

Table 2 :Correlation co-efficient (r) between x and y/x in Southern Highlands Province

Stratum	df	r	t	Status	c.v. of x %
Coffee	94	$\simeq$ 0	$\simeq$ 0	I	51.22
Non-Coffee	23	0.202	0.989	I	61.35
TOTAL	119	$\simeq$ 0	$\simeq$ 0	I	53.24

(Note : $t = \frac{r\sqrt{n-2}}{\sqrt{1-p^2}}$ with n - 2 d.f.

The value of r, as observed, is insignificant in all cases, indicating thereby a lack of correlation between number of households and trees per household, also there is a substantiate variation among household numbers (x).

The summary of the results in the above two tables is that the sampling scheme is efficient for "coffee" stratum, though not so for "non-coffee" one. In other words, while generally the scheme holds good for "coffee" stratum, a simpler scheme for "non-coffee" stratum could be worth consideration. In order to check whether this particular sampling scheme is more suited to any other related variable, an ancilliary variable, production of coffee, was considered. t - tests showed similar trends as number of trees ; furthermore, the comparison between these two sets of r-values (between x and y) by means of z-transformation did not produce any significant difference indicating that the scheme holds good for all important characteristics.

6. Test in North Solomons Province

As in case of Southern Highlands province, villages belonging to North Solomons province were also divided into two strata : those belonging to Community Governments that produced more than 4,000 bags of cocoa in 1980 were grouped into "large" cocoa-growing areas while the remaining grouped into "small" stratum. Similar sets of tests as in Southern Highlands province were carried out ; however, in view of their relative importance, both cocoa and coconut trees were considered for testing in all strata. The summary of the results is shown below (for testing of hypothesis (a)) :

Table 3 : Tests for Linearity of Regression in North Solomons Province

Stratum	d.f	a			b		
		Value	t	Status	Value	t	Status
I Cocoa							
Large	44	1,196	0.150	I	184,09	5.164	HS
Small	23	6,292	0.646	I	76,71	2.084	S
TOTAL	69	7,491	1.121	I	121,42	4.339	HS
II Coconut							
Large	44	1,130	0.426	I	20,80	1.754	I
Small	23	- 1,429	0.290	I	36,80	1.973	I
TOTAL	69	- 136	0.057	I	28,72	2.877	HS

(Note : S = Significant at 5 % level but Insignificant at 1 % level)

The summary shows that the scheme holds good for both cocoa and coconut in general ; however, in case of cocoa the scheme appears (not unexpectedly) more suitable for "large" stratum, while in case of coconut, division into strata according to cocoa production does not provide any linear relationship.

For testing of hypothesis (b) ; the correlation between x (population) and y/x (number of trees per head) was considered and the results are as follows :

Table 4 : Correlation co-efficient (between x and y/x)

Stratum	df	r	t	Status	c.v. of x %
I Cocoa					
Large	44	0.1069	0.713	I	46.12
Small	23	-0.1000	-0.482	I	48.19
TOTAL	69	-0.0259	-0.215	I	47.45
II Coconut					
Large	44	-0.0492	-0.326	I	46.12
Small	23	0.0366	0.175	I	48.19
TOTAL	69	-0.0054	-0.045	I	47.45

The value of r, as in Southern Highlands, is found to be insignificant for both strata and for both cocoa and coconuts, indicating thus a lack of correlation between population and trees per head ; at the same time, there appears to be a substantial variation among population of village (x).

The summary of the results shows that the general sampling scheme is suitable for both main variables, i.e., cocoa and coconut numbers of trees ; however, stratification of villages according to cocoa production does not seem to increase its efficiency.

7. Conclusion

On examination of results in Southern Highlands and North Solomons provinces, it can be conveniently said that two-stage sampling scheme as adopted (i.e., p.p.e.s. with replacement at the first stage and s.r.s. without replacement at the second) is quite effective for the recent series of provincial crop surveys though this involves quite a bit of preliminary work (e.g., probability assignment, selection, computation of raising factors) at the early stages ; it appears, however, a closer look is needed for meaningful stratification to achieve greater efficiency.

8. References

1. Koley, C. (1979). An appraisal of the sample survey of smallholding perennial crops in Papua New Guinea, 1975-76 ; Survey Statistician No. 2). International Association of Survey Statisticians.

2. Cochran, W. G. (1977). Sampling Techniques, 3rd Edition, John Wiley & Sons, Inc. Chap. 11, pp. 306-307.

3. Op. cit. Chap. 11, pp. 308.

4. Murthy, M. V. (1967). Sampling Theory and Methods. Statistical Publishing Society, India, Chap. 6 pp. 186, 189.

5. Cochran, W. G. (1963). Sampling Techniques, 2nd Edition, John Wiley & Sons, Inc. Chap 11, pp. 322.

6. Snedecor, G. W. & Cochran, W. G. (1967). Statistical Methods, Sixth Edition. Iowa State University Press, U.S.A. Chap. 6, pp. 153,,166-167.

Prepared by :

CHINMOY KOLEY, Rural Statistician
ZENUDDIN WALIJI, Senior Rural Statistics Officer
Policy & Planning Division, department of Agriculture & Livestock,
P O Box 417, KONEDOBU, Papua New Guinea, (March 1987).

NUMBER OF TREES, STANDARD ERROR AND COEFFICIENT OF VARIATION FOR FOUR IMPORTANT CROPS SURVEYED DURING SERIES OF CROP SURVEYS, 1979 - 84

PROVINCE	N	COCONUT				COCOA				ARABICA COFFEE				ROBUSTA COFFEE			
		n	Number of trees ('000)	S.E.	C.V. (%)	n	Number of trees ('000)	S.E.	C.V. (%)	n	Number of trees ('000)	S.E.	C.V. (%)	n	Number of trees ('000)	S.E.	C.V. (%)
MILNE BAY	439	99	3,076.0	283.9	9.2	63	191.4	69.8	36.5	19	340.8	81.9	24.0	66	100.2	32.0	31.9
CENTRAL	615	56	1,243.4	314.8	25.3	12	23.0	14.5	63.0	14	757.1	162.2	21.4	12	1.5	1.5	100.0
MOROBE	1,044	33	1,724.0	487.1	28.3	25	1,015.1	361.9	35.7	61	10,820.1	1,130.6	10.4	33	497.9	194.5	39.1
N. SOLOMONS	551	79	2,741.5	554.2	20.2	79	12,541.5	1,080.9	8.6	0	NOT	GROWN		0	NOT	GROWN	
MADANG	1,073	42	3,867.0	1,107.3	28.6	48	3,186.3	551.0	17.3	41	4,877.2	1,386.0	28.4	49	2,205.4	415.0	18.8
ENGA	648	0	NOT	GROWN		0	NOT	GROWN		29	4,773.3	890.6	18.7	0	NOT	GROWN	
E. NEW BRITAIN	392	39	3,918.1	635.0	16.2	39	6,399.2	804.6	12.6	0	NOT	GROWN		0	NOT	GROWN	
EAST SEPIK	977	14	268.9	140.1	52.1	37	2,250.3	400.2	17.8	0	NOT	GROWN		43	7,598.9	1,095.9	14.4
NEW IRELAND	369	55	3,702.0	306.0	8.3	55	1,264.8	207.1	16.4	0	NOT	GROWN		0	NOT	GROWN	
E. HIGHLANDS	987	0	NOT	GROWN		0	NOT	GROWN		58	29,156.0	2,992.9	10.3	0	NOT	GROWN	
S. HIGHLANDS	642	0	NOT	GROWN		0	NOT	GROWN		121	13,432.3	1,598.6	11.9	0	NOT	GROWN	

V - PAPERS ABSTRACTS

Binder, D.A., M. Gratton, M.A. Hidiroglou, S. Kumar, and J.N.K. Rao (1984). Analysis of categorical data from surveys with complex designs : Some Canadian experiences. Survey Methodology, 10, 141-156.

The standard statistical literature is rich with methods for performing data analysis for categorical data under certain parametric moledo, usually associated with multinaomial or Poisson distributions. This has led to a wide range of techniques for testing hypotheses and making other inferences about the underlying medel parameters. However, design-based inferences from samples derived from complex survey designs, such as stratified multi-stage designs cannot be based on the usual parametric model assuptions.

Recently, there have been a number of developments in the literature on how to take account of the survey design when analyzing categorical data sets. These analyses include goodness of fit tests, tests of independence, log-linear models and logistic regression analysis. Developments have concentrated on estimating of variances, deriving asymptotically chi-squared test statistics and obtaining null distributions and some approximations of "standard" Pearson Chi-squared tests and Likelihood Ratio tests.

In this paper, we review some of these developments and exemplify the methods as they apply to data from the Canada Health Survey and the Canadian Labour Force Survey. A brief discussion of availability of software for performing these analyses is also included.

Reprint requests should be sent to :

Dr. A. Binder
Statistics Canada
Jean Talon 4-D8
Tunney's Pasture
Ottawa K1A OT6
Canada

Authors : C.P. Quesenberry Jr and N.P. Jewell.

Title : Regression analysis based on stratified samples.

Journal : Biometrika (1986),73, No. 3, pp 605-614.

Regression analysis is often used in the analysis of survey data in situations where complex sampling designs are employed. This paper considers estimation of the p x 1 vector of parameters B in the assumed population model $E(Y|X) = X B$ under stratified sampling on the dependent variable Y. Also considered is the case of stratified sampling on a design variable Z correlated with the dependent variable but not included as an independent variable. Under such sampling schemes, the ordinary least squares estimator is generally asymptotically biased. One approach to estimation is to make assumptions as to the form of the joint distribution of X, Y and Z, and estimate by the maximum likelihood method. Previous authors have assumed a multivariate Gaussian distribution. This paper focuses on estimators that require minimal assumptions on this joint distribution.

Currently, there are two methods that do not use parametric assumptions. The first is a weighted least squares approach where the weight associated with each observation is inversely proportional to its probability of selection. Under mild moment conditions on the conditional distributions of X and Y given Z, this probability weighted estimator has been shown to be asymptotically unbiased, but it can be extremely inefficient. In an attempt to improve on the efficiency of the probability weighted estimator and to maintain freedom from any restrictive distributional assumptions, Jewell (Least squares regression with data arising from stratified samples of the dependent variable ; Biometrika (1985) 72, 11-21) proposed an iterative least squares estimator for the case of stratified sampling directly on the dependent variable. His estimator has been shown to be consistent ans asymptotically Gaussian, and showed good efficiency in his preliminary simulation study. Here, in a more extensive Monte Carlo study, it is shown that Jewell's (1985) estimator can also be inefficient under some sampling conditions.

This paper introduces a new iterative least squares method. As shown in a simulation study, it is generally more efficient than both the probability weighted estimator and Jewell's (1985) estimator. The proposed estimator is similar to Jewell's in that it involves iteratively transforming the data and applying the ordinary least squares method on the transformed data set. The main difference is that where Jewell (1985) transforms the data by applying multiplicative adjustments to the dependent variable, the new method applies additive adjustments. The approach also generalizes to the case of sampling on a design variable Z correlated with the dependent variable Y. Estimation of the variance of the nex estimator is considered and a preliminary simulation study is presented which shows that the proposed variance estimator performs well in cases of moderate disparity among the selection probabilities but appears to be unstable in some of the more extreme cases studied. Further research on variance estimation is ongoing.

Reprint requests should be sent to :

Charles P. Quesenberry. Jr.
The Permanente Medical Group Inc.
Division of Research
3451 Piedmont Avenue
Oakland, California 94611-5463

D.R. Bellhouse
A Review of Optimal Designs in Survey Sampling
Canadian Journal of Statistic, 1984, 12 : 53-65

In many sampling problems attention has been focussed on optimal (for example, best linear unbiased) estimation of the finite population mean. The optimal estimator is usually obtained within a wide class of sampling designs. By comparison the problem of optimal selection of the sampling design has not been as widely addressed. This paper attempts to bring together in a literature review several diverse topics but with the common theme centered around the choice of the sampling design.

On assuming various linear regression superpopulation models with uncorrelated errors for the finite population units, there are two routes that one can take in the selection of an estimator of the mean : make the choice of the estimator from among design-unbiased estimators or from among model-unbiased estimators. If the route of model-unbiased estimation is followed then the optimal sampling design reduces to various methods of purposive selection. On the other hand if design-unbiased estimators are used the optimal design if often any sampling design which has inclusion probability of the j^{th} unit $\pi_j \propto \sigma_j$, where σ_j^2 is the model variance of the measurement associated with the j^{th} unit. When the uncorrelated error structure is replaced by autocorrelated errors and the auxiliary variables are removed from the regression equation, then the optimal choice of the design reduces to a unique choice rather than a class of designs. This optimal design is systematic sampling of variations of it, depending on the choice of the autocorrelation function and whether or not design-unbiased estimation is required.

Two other topics are reviewed in the paper : minimax sampling designs and controlled sampling. The main result reviewed under the former topic is that simple random sampling without replacement is the minimax sampling design even using general assumptions about the population. In controlled sampling the object is to obtain sampling designs for which the inclusion probabilities and joint inclusion probabilities are the same as simple random sampling without replacement but for which there are fewer than $\binom{N}{n}$ possible samples. The solution to this problem requires results in the theory of balanced incomplete block designs.

Reprints requests should be sent to :

Dr. David Bellhouse
University of Western Ontario
Dept. of Statistical and Actuarial Sciences
London, NGA 5B9
Canada

Greg J. Duncan and Graham Kalton : issues of Design and Analysis of Surveys Across Time
(International Statistical Review (1987), 55, 97-117)

Both the composition and characteristics of populations change over time, raising a host of issues for the design and analysis of surveys that are concerned with populations over time. Since resolution of these issues depends on the survey objectives, we begin with a listing of possible objectives, including : estimating population parameters at distinct time points, during distinct intervals or averaged across periods of time ; measuring net change at the aggregate level, or measuring gross change, average change or instability at the individual level ; measuring the frequency, timing and duration of events ; and cumulating samoles of rare populations over time.

We next show how these objectives are met by four survey designs : repeated surveys (that employ fresh samples at each time point) ; panel surveys (that take repeated measurements on the same elements) ; rotating panel surveys (that take repeated measurements over a restricted interval) ; and split panel surveys (that combined panel surveys with repeated or rotating panel surveys). Repeated surveys automatically take population changes into account but cannot estimate gross change or other components of individual change. Panel surveys are well-suited for analysis of individual change. However, they need a mechanism for taking population changes into account if estimates are desired for populations at time points subsequent to the initial sampling point. Rotating panels can provide estimates of individual change over the course of the rotation period and handle population changes with new rotation groups. Split panel surveys accomplish many of the purposes of both repeated and panel surveys.

We next consider a set of potential problems of varous survey designs, including initial wave nonresponse bias, subsequent wave nonresponse, and conditioning bias. We also consider some aspects of panel designs that may improve data quality, including increasing respondent motivation, shortening the recall period for retrospective studies and using interview information from the previous wave for bounded recall. The four survey designs are evaluated with respect to these potential problems and benefits.

Nonresponse is inevitable in panel survey designs, although a number of panel projects have attained surprisingly high response rates. Various methods for minimizing nonresponse in panel surveys are reviewed, as are methods for compensating for nonresponse, including imputation and weighting.

Finally we review some recent methods of analyzing longitudinal survey data, including the description of gross change, event history models, fixed-effect models that difference out unobserved individual-specific factors, and error-component models.

Reprint requests should be sent to : Greg J. Duncan or Graham Kalton, Survey Research Center, University of Michigan, Ann Arbor, MI. 48106-1248, USA.

Phillip S. Kott : nonresponse in a periodic sample survey
Journal of Business and Economic Statistics, April, 1987, Vol. 5, No.2, pp.287-293.

A parametric model is introduced to analyze alternative methods of imputation for nonresponse in a periodic sample survey of a continuous variable. Although it is assumed that nonresponse is ignorable (not related to the variable being surveyed), the parametric model does not require that response propensities be random. On the other hand, it is argued that for certain imputation strategies a quasi-random response model can be invoked to provide some protection against parametric model failure.

If the population units follow the parametric model, then imputing for a missing variable by updating the nonrespondent's historical value with a "ratio-of-identicals" of the respondents is shown to be better than imputing with a respondent mean in all but one very rare circumstance (a very large negative correlation between each unit's historical value and the difference between its present and updated historical values). Although the exact specification of an historical value is left vague for theoretical purposes, it is noted that in practice the historical value of a variable is most commonly its last reported value.

A test is developed for comparing alternative formulations of the historical values. An analysis of monthly gasoline sales data consisting of a large number of cells reveals that updating last month's value is better than imputing with the respondent cell mean (a degenerate case of updated historical imputation) in a great majority of cases. Furthermore, significantly more often than not there are additional gains to be made (in terms of reducing model mean squared error) from using, as the historical value, an exponentially smoothed version of past values which gives greatest weight to last month's value.

This article provides a formal theoretical justification for a popular ad hoc imputation technique - the ratio-of-identicals method. By so doing, it begins to develop the analytical tools necessary for assessing the power and limitations of this approach in many surveys to which it either is or should be applied.

Reprint requests should be sent to :

Dr. Phillip S. Kott
Energy Information Administration
1000 Independence Avenue, S.W.
Washington D.C. 20585
USA

Ranjit de Mel :

Feasibility of a Two Frame Approach to the Monthly Population Survey

Statistical Services Papers, Australian Bureau of Statistics (1987, Vol 5, No.1).
Available from D.G. Steel (Editor), c/- PO Box 10, Belconnen, 2616

An investigation of the operational features of telephone interviewing for the Monthly Population Survey (MPS) recommended a methodology in which the initial interview was made in the field ans, for those whor agreed, subsequent interviews be made by telephone. In this methodology the sample enumerated byu telephone is still tied to the area sample and so the question arises ; can significant savings be achieved by separating the telephone and area sample ?

The paper examines the possible cost savings of using a telephone ownership list (frame B) in conjunction with the area frame (frame A) as a means of reducing the cost of the MPS.

The paper concludes that the cost savings are minima and, in wiew of additional implementation and operational complexities, should not be pursued.

VI - QUESTIONS/ANSWERS

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

Question : Several of your Q/A columns deal with computations for variances and for sampling error functions based on variances. But are these variances and sampling errors all that important, when there are so many other sources of errors and variation in survey results ? A) Are sampling errors sufficient, when the "total error" may include even greater sources of errors and biases due to measurements and nonresponses ? B) Are sampling errors necessary when dealing with data from censuses or administrative records ?

Answer : Indeed I have often heard this important question (A above) stated in different ways, as for example : "Survey results are subject to measurement biases that are generally even larger than sampling errors. Therefore, computing and presenting sampling errors only - instead of "total errors" - is not only insufficient, but it is even misleading. Sampling errors are not worth all the cost and efforts needed to produce them". Here are my answers to these objections.

1) Sampling errors can be computed from the survey data themselves ; although some care is needed to make the design "measurable" (See Q/A in SS 13 and Kish, 1987, 7.1E) and modest extra costs for computing them. On the other hand, nonsampling errors generally require data collected from outside (beyond) the survey data ; those, and especially the estimates (if any) for measurement biases need considerable (or great) extra costs and efforts. Sampling errors can be computed with modern programs for a few hundreds or thousands of dollars, less than 1 percent of some projects' costs. If you cannot do that, look for "Alternatives for Sampling Errors for Complex Samples" [Kish 1987, 7.1B].

2) Sampling statisticians can compute sampling errors, but measurements of nonsampling errors and biases need the knowledge of experts in the specific subject matter.

3) Nonsampling errors come from the multitude of all survey operations, from all aspects of data collection and processing. The systematic (average) biases for each of these may be even greater, more important and more elusive and difficult to measure than its variable errors.

4) For the above reasons, measuring all or most nonsampling errors and biases is and will remain impossible for any single survey. Thus measuring "total errors" is not reasonable even as a distant, eventual goal. Knowledge about nonsampling errors and biases must be cumulated over entire fields, and field by field, rather than measured and computed for single surveys, which can and should be done for sampling errors.

5) But it would be wrong to refrain from computing and presenting sampling errors, just because nonsampling errors cannot be presented equally simply and concisely. The public should be warned (and they often are) that beyond sampling errors, survey statistics are also subject to the systematic biases that complete

censuses would also suffer. That sampling errors are greater for subclasses, and much greater for small subclasses, than for the entire sample, should also be shown to thereaders of survey results (See Q/A in SS 16 and 17).

6) Most important : Whereas measurement biases are often greater than sampling errors for overall statistics, variances usually dominate the errors for subclasses ("crossclasses"), comparisons, and analytical statistics. That is (or should be) the reason for the sizes of the samples. It is common practice foranalysts to exploit survey data to the limits permitted by the sample sizes - ans often, alas, even well beyond those limits. This figure [Kish 1987, 2.4.1] illustrates such common situations.

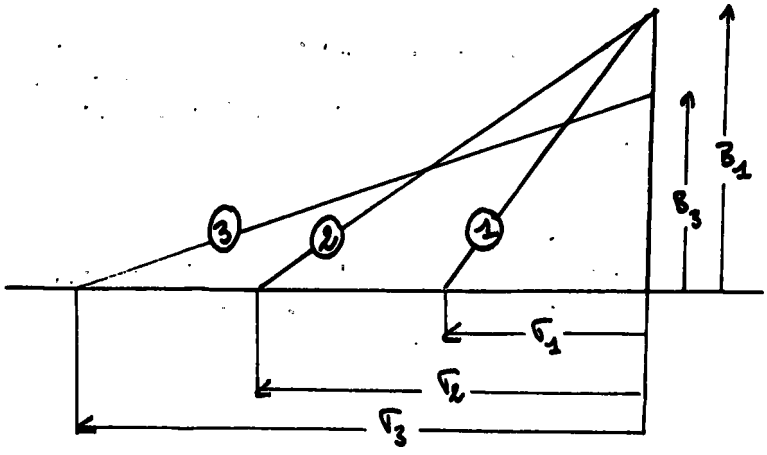

Figure : variable errors (σ) and biases (B) in root mean square errors (RMSE).

The bases represent sampling errors and other variable errors (σ). For example, σ_1 may be the ste ($\bar{y}_1$) for the mean $\bar{y}$ of the entire sample and $\sigma 2$ may be a larger ste ($\bar{y}_2$) for a subclass mean, and σ_3, may be the ste ($y_a - y_b$) for the difference between two subclass means.

The heights represent biases (B) and the hypotenuse denotes the RSME = $\sqrt{(\sigma^2 + B^2)}$; (see 7.2 F).(1) For the entire sample the bias B_1 may be large compared with the variable error σ_1, thus taking larger samples would not decrease the $RMSE_1$ by much.(2) However, with the same bias B_1, but with a smaller sample in the subclass, the ratio changes and the σ_2 dominates the $RMSE_2$; and this is not much larger than for (1) despite a much smaller sample (3). Furthermore, for the difference of means, the net bias B_3 may be much smaller ; so that even with a larger σ_3, the $RMSE_3$ for the difference is but little greater than $RMSE_2$. This drastic change in the bias ration B/σ tends to appear not only for differences between subclasses within the same sample, but also for differences between repeated surveys.

Question B refers to structural variations whereas Question A concerned measurement errors. It is best to separate these two alternative sources of variation, both competing with the sampling errors. When demographers and economists use complete censuses, registers, and administrative records, are those statistics free from stochastic variations ? I believe that analyses are aimed beyond the "target" or "frame" populations to some "inferential" population(s) (See Q/A in S.S.15). Thus the census populations exist in some matrix of variations both in space and time. To deny the relevance of those variations would mean that any small difference between two periods (between successive days or minutes) or between two places (countries or blocks) would need and deserve causal explanations. [Brillinger

1986]. Let me add that I believe the variations of actual populations from "superpopulations" to be usually greater and more complex than Bernouilli or Poisson models would indicate.

Brillinger DR [1986], Natural variability of vital rates, Biometrics 42, 693-731.

Kish L [1987], Statistical Design for Research, New York : John Wiley and Sons.

Addition to Q/A in SS 18 :

A vigilant reader points out that point g) in the answer dealing with Collapsing strata of PSUs would benefit from reference to a new article : Rust K and Kalton G [1987], Strategies for collapsing strata for variance estimation, Journal of Official Statistics, 3, 69-81.

Also to an older one :

Du Mouchel WH, Govindarajulu Z, and Rothman E [1973], A note on estimating the variance in stratified sampling, Canadian Journal of Statistics, 1(2), 267-274.