USE OF INTERPENETRATING SAMPLES IN DEMOGRAPHIC STUDIES

By RANJAN KUMAR SOM

*Indian Statistical Institute**

SUMMARY. The technique of interpenetrating (net-works) of sub-samples of Mahalanobis is a powerful tool in : (a) examining the enumerator bias and computing the sampling error by the analysis of variance; (b) supplying margins of uncertainty of the estimators; (c) giving an "error area" for a curve and the "separation" of the curves for two sectors or time periods by the method of fractile graphical analysis; (d) providing control in collection, processing, and analysis; and (e) supplying advanced estimates on basis of part-samples. The use of this technique in the field of demography has been illustrated with numerical examples.

INTRODUCTION

1. The technique of interpenetrating (net-works) of sub-samples (IPNS) consists of dividing the total number of units in an enquiry (either on a complete enumeration or on a sample basis) into a number of parallel, random groups and assigning one such group to each of the factors causing variation (e.g., enumerators, variation in field schedules and different methods of collection). It was introduced by P. C. Mahalanobis in 1937 in Crop Surveys in Bengal (Mahalanobis, 1938-41; 1946).

2. The technique of IPNS has been used in a number of sample surveys in India including the National Sample Survey (1950–57), the Demographic Sample Surveys in Southern Rhodesia (1948 and 1953-55), and the Philippine Statistical Survey of Households (1957).

DETECTING ENUMERATOR BIAS

3. A simple tabulation may often reveal differences in the estimates obtained by different enumerators covering the same universe. The arrangement of the total units into a number of sub-samples enables one to test by the method of Fisherian analysis of variance the differential effects of enumerators (parties) and of other factors. Cochran (1953) has given a theoretical model when a simple random of n units is divided at random into k groups of units, each group containing $m = n/k$ units, and surveyed by one enumerator (or party of enumerators) each.[1]

---

*At present a member of the Secretariat of the United Nations at the Economic Commission for Africa, Addis Ababa. The views presented are the author's own.

[1] For the development of mathematical models for response errors see Hansen, Hurwitz, Marks and Mauldin (1951), Sukhatme and Seth (1952), Hansen, Hurwitz and Nadow (1953), Sukhatme (1953), and Hanson, Hurwitz, and Borshad (1961). It is also possible, by the use of the Studentized smallest χ^2, to test the significance of the smallest of the differences between two enumerators: a significant smallest variance ratio would lead one to suspect collusion between the two enumerators (Ramachandran, 1958).

4. A net bias common to all enumerators (or to all the components of any factor causing variation) admittedly passes undetected by this technique.[2] Such biases have to be controlled by improvement in survey methodology, comparison with independently obtained estimates, and adoption of special analytical techniques.

5. Previously unpublished examples of the use of the analysis of variance to test the effect of enumerators and of other factors are given below.

6. *Demographic sample surveys in Southern Rhodesia.* In the 1948 Demographic Sample Survey of the African Population of Southern Rhodesia interpenetrating samples were used to provide a check on the accuracy of the field work, each field officer being allocated the whole of a sub-sample in three successive districts, so that his work could be compared with that of at least three other officers, by testing whether the two or three sample concerned had been drawn from the sample population. "In practice the tests did not detect any significant differences with regard to total population, only three out of over seventy sub-samples were discarded for birth and infant mortality rates, but eighteen were thrown out in calculating the death rates."[3]

7. In the 1953 Demographic Sample Survey of the Indigenous African Population of Southern Rhodesia, interpenetrating samples were used to determine the sampling errors. However, from the basic data, we have constructed in Table 1 the analysis of variance of birth and death data relating to six districts in 1953. The ten sub-samples adopted were not statistically significant, showing that the survey was under statistical control; the six districts, however, showed significant variation in regard to birth and death data.

TABLE 1. ANALYSIS OF VARIANCE OF (ESTIMATED) NUMBER OF BIRTHS AND DEATHS DEMOGRAPHIC SAMPLE SURVEY OF THE INDIGENOUS AFRICAN POPULATION OF SOUTHERN RHODESIA, 1953[a]

(6 districts × 10 sub-samples)

source	d.f.	births		deaths	
		m.s.	F-ratio	m.s.	F-ratio
(1)	(2)	(3)	(4)	(5)	(6)
districts	5	123103.95	15.48*	48184.43	19.45*
sub-samples	9	7531.81	1.06(r)	2050.44	1.21(r)
error	45	7950.03	—	2476.77	—
total	59	17645.04	—	6285.27	—

*Significant at 0.1% level. (r) Indicates error/sub-sample in F-ratio.

[a]Basic data supplied by courtesy of the Director of Census and Statistics, Statistical Office, Rhodesia.

8. *Indian National Sample Survey,* 1955. In the Indian NSS 9-th round, 1955, the sample design was two-stage stratified with two "parties" of enumerators in a State, each party working in two time periods (the latter consisting of three months each), giving four estimates of any characteristic under study. Som and

[2]Mahalanobis (1946): Foreword to reprint, (1960a).

[3]Personal communication from Dr. C.A.L. Myburgh, Director, Statistical Office, Rhodesia.

Mukherjee (1962) have analysed the data of rural Uttar Pradesh for strata having 4, 8 and 12 sample villages each, the data for each village being taken as the basic observation. The analysis of variance was done for four demographic characteristics and the results are summarized in Table 2 which gives the distribution of the theoretical and the observed party/error F-ratios for strata having 4 and 8 sample villages each; the values of χ^2 calculated to test the goodness of fit and the associated probabilities are also shown in this table. Were the χ^2 's independent, it would have been possible to pool these probabilities by the $P-\lambda$ test to arrive at an overall probability : such a pooled probability comes to 0.37 and 0.64 respectively. The χ^2's are not, in fact, independent and the observed values would over-estimate the probabilities. Since, however, the probabilities are quite high, we could reasonably infer that the observed frequency distributions conform fairly well to the theoretical distributions, that is, that there was no significant party-differences and that the survey was under statistical control.[4]

TABLE 2. DISTRIBUTION OF THEORETICAL AND OBSERVED PARTY-ERROR F-RATIOS AND ASSOCIATED VALUES OF χ^2 : NSS 9-TH ROUND, RURAL, UTTAR PRADESH, 1955

range of F-ratio (from the right-hand tail-end)	theoretical frequency	observed frequency corresponding to			
		sex-ratio	percentage of population economically active	percentage of literate persons	percentage of persons aged 50 above
(0.1)	(0.2)	(1)	(2)	(3)	(4)
For party/error F-ratio (1, 1 d.f.) in strata with 4 sample villages each					
—F60%	10	7	11	13	9
F60%—F20%	10	14	13	7	10
F20%—	5	4	1	5	6
total	25	25	25	25	25
χ^2(2 d.f.)		2.70	4.20	1.60	0.30
P		0.26	0.13	0.42	0.86
For party/error F-ratio (1, 4 d.f.) in strata with 8 sample villages each					
—F40%	10.8	13	9	10	12
F40%—	7.2	5	9	8	6
Total	18	18	18	18	18
χ^2(1 d.f.)		1.12	0.75	0.15	0.33
P		0.29	0.40	0.70	0.58

[4] Further examples of analysis of variance of demographic variables are given in Mahalanobis(1946) and Indian NSS Report No. 48, Appendix 2.

MARGIN OF UNCERTAINTY

9. In sample designs, it is customary to calculate the error from the estimates obtained at the first stage (selected with replacement). Such an error includes also the differential effect of the enumerators. In sample designs with interpenetrating sub-samples also, errors if calculated in this manner would include the differential effects of the enumerators.

10. In the model for IPNS with a simple random sampling adopted by Cochran (1953), an unbiased estimate of the sample variance of the sample mean $\bar{y}$ was obtained as

$$v(\bar{y}) = \frac{\sum_i (\bar{y}_i-\bar{y})^2}{k(k-1)}$$

where $\bar{y}_i = \sum_{j=1}^{m} y_{ij}/m$ $(i = 1, 2, \ldots k)$ is the mean for the i-th group (party), and $\bar{y} = \sum_i \sum_j y_{ij}/mk$, the grand mean.

11. In multi-stage stratified designs, most commonly met with, the margin of uncertainty can be built up from the two (or more) IPNS, which provide independent, equally valid estimates at the stratum level. This was done in the 1953-55 Demographic Sample Survey of Southern Rhodesia and is being followed in the Indian NSS.[5]

12. Thus if y_{hi} is an (unbiased) estimate for an aggregate for the h-th stratum from the i-th IPNS $(i = 1, 2, \ldots, k)$, the estimated value at the stratum level from all IPNS would be :

$$\bar{y}_{ho} = \sum_i y_{hi}/k$$

and the estimated aggregate over all strata

from the i-th IPNS : $y_{oi} = \sum_h y_{hi}$ ... (1)

from all IPNS: $\bar{y}_{oo} = \sum_h \bar{y}_{ho}$ ... (2)

$= \sum_i y_{oi}/k.$ ... (3)

The estimated variance of $\bar{y}_{oo}$ calculated from formula (2) is :

$$v(\bar{y}_{oo}) = \sum_h v(\bar{y}_{ho}) = \frac{\sum_h \sum_i (y_{hi}-\bar{y}_{ho})^2}{k(k-1)} \quad \ldots (4)$$

and the estimated variance of $\bar{y}_{oo}$ calculated from formula (3) is, since the estimates obtained from the k IPNS are independent

$$v'(\bar{y}_{oo}) = \frac{\sum_i (y_{oi}-\bar{y}_{oo})^2}{k(k-1)} = \frac{\sum_i [\sum_h (y_{hi}-\bar{y}_{ho})]^2}{k(k-1)} \quad \ldots (5)$$

[5] See, for example, NSS Report No. 48, Appendix 2 ; and No. 49, Appendix 2.

Both the estimated variances are unbiased.[6] The variance $v(\bar{y}_{00})$ will have effective degrees of freedom lying between $(k-1)$ and $L(k-1)$ where L is the number of strata (Satterthwaite, 1946; Cochran, 1953); the variance $v'(\bar{y}_{00})$, is much simpler to compute, although being based on a smaller number of degrees of freedom ($= k-1$), will be less efficient.[7] With two sub-samples, for example, the estimated standard error of the estimated $\bar{y}_{00}$ from formula (4) is :

$$s(\bar{y}_{00}) = \tfrac{1}{2}\sqrt{\Sigma\,(y_{h1}-y_{h2})^2}$$

and from formula (5) is simply : $s'(\bar{y}_{00}) = \tfrac{1}{2}\,|\,y_{01}-y_{02}\,|$ where y_{h1} and y_{h2} are the stratum estimates and y_{01} and y_{02} the estimated over all strata from the two sub-samples. Both these standard errors are easy to compute.

13. Table 3 shows in row 1 the estimated variances of the birth and death rates in the Kerala State in India, obtained from the Indian NSS, 14-th round, 1958–59: these were calculated from the variance and (corresponding covariance) formulae of the type (4) from differences of the estimates from the two parties at the stratum level. The error variances, obtained from an analysis of variance done on the data and shown in row 2 of the table, are, as expected, smaller than the corresponding variances obtained earlier. The variances obtained by ignoring stratification are also shown in row 3 of the table.

TABLE 3. ESTIMATES, ERRORS, AND PERCENTAGE ERRORS OF BIRTH AND DEATH RATES FOR THE KERALA STATE : INDIAN NSS, 14-TH ROUND, 1958-59

type of estimate	birth rate/1000			death rate/1000		
	estimate	error	percentage error	estimate	error	percentage error
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1. ratio of simple unbiased aggregate of the variate to that of population	37.26	2.28[b]	6.12	11.64	0.69[b]	5.93
2. simple average of the rates of the 72 sample villages	36.53	0.96[c]	2.63	11.16	0.66[c]	5.91
3. simple average of the rates of the 72 sample villages	36.53	1.07[d]	2.93	11.16	0.65[d]	5.82

[a] NSS Report No. 48, Appendix 2.
[b] Obtained from the differences of the two parties at the stratum level.
[c] From the interaction "stratum × party × time" in analysis of variance.
[d] Ignoring stratification and calculating errors for two parties separately.

[6] See also Lahiri (1954), United Nations (1960), and Zarkovich (1963).

[7] But on this point see Koop (1960), who has also given an illustration relating to price index numbers collected by sampling methods, where the computation of the variance of the classical type is not normally feasible, but the technique of IPNS would supply variances in the above forms.

14. Mahalanobis has shown a simple non-parametric way of expressing the margin of uncertainty of the estimators from the IPNS. Each of the independent, interpenetrating sub-samples provide an independent and equally valid estimator, not only for the principal variates but also for the derived functions, and directly give, therefore, an estimate of the margin of uncertainty involved in the estimator. This is an advantage of particular relevance to surveys in underdeveloped countries where suitable machines for the calculation of variances (not only for a principal variate, but a large number of variates and their derived functions) are not in plentiful supply. The probability that the median of the distribution of the sample estimates (which is assumed to be symmetrical) which will converge to the population mean in large samples, will be contained by the range of k estimates obtained from the k independent, interpenetrating sub-samples is $1-\left(\frac{1}{2}\right)^{k-1}$: with two sub-sample estimates for example, the range between the two estimates provides 50 per cent margin of uncertainty for any derived functions worked out from these two samples. The derivation of such limits, besides being simple, involves a minimum of assumptions regarding the sampling distribution of the estimate.

15. *Illustrations of margins of uncertainty.* Table 4 illustrates the use of IPNS in setting margins of uncertainty for some selected demographic variables. The 50 per cent margins of uncertainty provided by the two independent, interpenetrating sub-samples are quite small and reasonable.

TABLE 4. MARGINS OF UNCERTAINTY FOR SOME SELECTED DEMOGRAPHIC VARIABLES

item	estimates			percentage deviation[a]	sample size
	sub-sample 1	sub-sample 2	combined		
(1)	(2)	(3)	(4)	(5)	(6)
1. birth rate 1000 persons, rural India 1958-59[b]	38.50	38.02	38.26	1.25	234,344 households
2. death rate/1000 persons rural India, 1959-60[c]	14.87	15.29	15.08	2.79	240,372 households
3. infant mortality rate/1000 live births, rural West Bengal, 1955[d]	173.00	104.26	109.16	5.52	5,539 live births
4. average age at marriage of wife (years), urban India[e]	15.0	14.6	14.8	2.70	6,175 couples
5. incidence rate for sickness/1000 persons, urban India, 1953-54[f]	46.51	44.57	45.58	4.20	1,720 households
6. economic activity rate/100 persons, rural India, 1953-54[g]	44.70	45.14	44.92	0.98	8,235 households

[a] 100 + sample 1 estimate—sample 2 estimate | ÷ combined estimate.
[b] NSS Report No. 48.
[c] NSS Draft No. 110.
[d] Poti *et al.* (1959).
[e] Das Gupta *et al.* (1955).
[f] NSS Report No. 49.
[g] NSS Report No. 14.

16. By the method of IPNS adopted in the Indian NSS round, it is possible to calculate separate sectoral estimates for any desired breakdowns by regions, social, or economic-classifications; but the margin of uncertainty would naturally increase with decreasing size of the sample breakdowns for any round. On the other hand, if the survey methods remain basically the same from one round to another, results from the different rounds can be pooled together to supply estimates averaged over the time-periods covered by these rounds; and such pooled data would be based on a much larger size of sample than that for any single round. Further, as Mahalanobis and Das Gupta (1954) have pointed out, two independent estimates would continue to be provided in whatever way the data for different rounds are pooled together or broken down. One important pre-condition in such pooling up is that the same parties should be labelled by the same independent sample number over the different rounds. This is being done in the Indian NSS from the 14-th round (1958–59).

17. When as is usually the case, one is more interested in estimates of ratios rather than of aggregates, there is an added advantage in having IPNS, in that the bias, generally inherent in ratio estimates, can be reduced on basis of two IPNS (Murthy and Nanjamma, 1959).

4. Fractile analysis

18. The method of fractile graphical analysis, introduced by Mahalanobis (1960b) in econometric studies is dependent on IPNS and has applications in demographic studies.

19. Fractile analysis of the total children born per couple by duration of marriage has been shown in Table 5 for the rural sector, and in Table 6 for the urban sector from NSS 3rd round, August-November 1951. To get an equal number of sample households in each of the 20 marriage duration groups adopted, the sample households were sub-sampled with probability proportional to the multipliers, thus rendering a self-weighting design. The couples in these households were then arranged in ascending order of magnitude of the duration of marriage for each of the sub-samples and the combined sample separately and the children born per couple in the "vigintile" (five percentile) groups calculated and plotted in the Figure below. The area between the curves for the two subsamples would give the "error" of the combined curve, and the "separation" between the rural and urban sectors could be measured by the divergence of the combined curves for the two sectors. The "error" of the urban curve is, as expected, seen to be much higher than that for the rural : the "separation" is, however, comparatively less.

20. *Two-way fractile analysis.* A two-way fractile analysis of the data on children per couple has been made in Tables 7 and 8 for the rural and urban sectors respectively. For any sample, the couples were first arranged according to the ascending order of duration of marriage and secondly according to the per capita household expenditure. This two-way arrangement could easily be done in the Tabulating Machines when a serial number according to the order of duration and another for that on per capita expenditure are punched. These results cannot, however, be presented graphically.

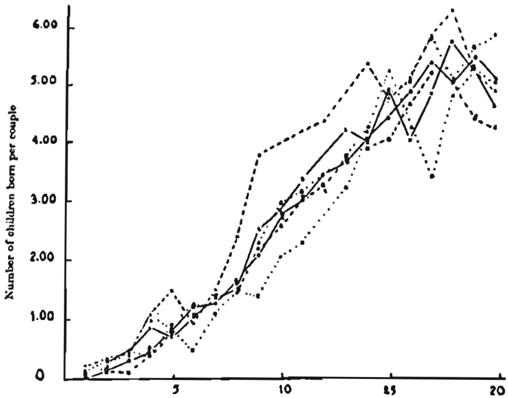

Fig. 1. Fractile group of marriage duration

Indian National Sample Survey : 3rd round : August-September, 1951

Rural		Urban	
sub-sample 1 :	o · · · o · · · o	sub-sample 1 :	x · · · x · · · x
sub-sample 2 :	o – – – o – – – o	sub-sample 2 :	x – – – x – – – x
combined :	o —— o —— o	combined :	x —— x —— x

IPNS in statistical analysis

21. The interpenetration of samples adopted in the Indian NSS was still recently only field-deep. The sample units in the field are arranged in interpenetrating sub-samples, each covered by a separate enumerator. But they lose their identity when they are processed, tabulated and analysed. To provide operational control at all the subsequent stages (Mahalanobis, 1946), it is necessary that interpenetration be also introduced at the post-field levels of processing, tabulation and analysis. A simple, live example may be of interest. In calculating the ratio of two estimates, it was found that the ratio was coming out inordinately low : this was traced to a correspondingly high figure for the denominator, due to a fault in the Calculating Punch Machine (the Multiplier) utilized for giving unbiased estimates for the numerator and denominator separately. If the calculation were arranged in two (or more) Calculating Punch Machines (according to the number of interpenetrating samples), the mistake would have been at once evident and would have saved much wastage.

IPNS in epidemiological surveys

22. Interpenetration can also be extended further and deeper than the field especially in epidemiological surveys where this technique would require two or more independent teams working in the field and the laboratory within the frame work of

the sample design (Cochran, 1953). It is also possible sometimes to divide one field team into two or more self-contained units. The independence and interpenetration of the sample units may be carried out to the extent possible.

23. Coming to a specific epidemiological field, in a tuberculosis survey this technique would require two (or more) independent teams collecting the samples, reading the X-ray films, and making bacteriological tests, the sample-units being arranged interpenetratingly within each domain of study : the above technique, it will be realized, is different from the independent reading of all X-ray films by more than one reader without independent, interpenetrating samples. In the latter situation, the individuals on whom the X-ray readings are made remain invariant for the readers and the independent readings would give an estimate of the bias of the readers only. With interpenetrating samples, on the other hand, the individuals in each of the independent samples would be different and the difference between the sample estimates would consist of both the sampling errors and the non-sampling errors, arising from different sources of variations—the bias of the informants, the bias of the observers, the bias in the methods of work, etc. In a tuberculosis survey, where the bias of the reader of the X-ray films is important enough to be reckoned in the sample design, independent readings of the films by more than one reader can easily be fitted in the framework of an interpenetrating sample design.

Advanced estimates from part-samples

24. Provisional, advanced estimates can, if required, be obtained on earlier processing of one (or more) of the IPNS (Mahalanobis, 1946; Aggarwal, 1963). The IPNS can also act as an insurance against unforeseen contingencies in the field (high turnover of, or accident to, investigators) or in the processing, as alternative estimates would then be available for any domain of study.

IPNS and supervision of field work

25. While supervisory checks are indispensable for securing adequate control over field work of the enumerators, the technique of IPNS, by detecting gross errors, can help in better supervision in the field work (as also in statistical tabulation). Moreover, if supervision can be arranged on a probability basis, the results can be utilized to improve those obtained by the primary enumerators (Sukhatme, 1953): the units surveyed by the supervisors can then be considered to form one interpenetrating sub-sample.

IPNS and additional travel cost

26. The technique of IPNS involves additional travel cost of the investigators when the same field area is covered independently by the different parties of investigators. Mokashi (1950) has given a formula for the loss of information due to this as a function of the cost of journey and other time per unit area. Mahalanobis (1950) observed that IPNS do not increase the survey budget (including planning, processing, and analysis) more than 10 per cent at an average; and a preliminary analysis of one round of the NSS data showed the loss of information as 3½ per cent of the total cost.

TABLE 5. TOTAL NUMBER OF CHILDREN BORN PER COUPLE, BY FRACTILE GROUPS OF MARRIAGE DURATION : NSS 3RD ROUND : ALL-INDIA (RURAL) AUGUST-NOVEMBER 1951

(909 villages : 1,800 households)

vigintiles*	percentile limits of marriage duration of couples (in years)			total number of children born per couple		
	sample 1	sample 2	combined	sample 1	sample 2	combined
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	01	01	01	0.01	0.05	0.04
2	02	02	02	0.20	0.12	0.15
3	04	03	04	0.48	0.10	0.30
4	05	05	05	0.39	0.54	0.51
5	06	06	06	0.93	0.78	0.87
6	07	07	07	1.23	1.10	1.22
7	09	09	09	1.33	1.41	1.29
8	11	10	10	1.62	1.48	1.65
9	13	13	13	2.18	2.29	2.15
10	15	15	15	2.60	2.80	2.80
11	16	16	16	3.20	3.05	3.05
12	18	18	18	3.60	3.33	3.49
13	20	20	20	3.80	3.80	3.76
14	23	22	22	4.29	3.95	4.09
15	25	25	25	4.90	4.09	4.46
16	28	28	28	5.06	4.71	4.95
17	30	30	30	5.85	5.25	5.43
18	35	36	35	5.16	5.12	5.15
19	40	43	42	5.70	5.43	5.56
20	66	65	66	5.88	4.33	5.15

*The word "vigintile" has been coined from a Latin root, meaning one-twentieth part.

TABLE 6. TOTAL NUMBER OF CHILDREN BORN PER COUPLE, BY FRACTILE GROUPS OF MARRIAGE DURATION : NSS 3RD ROUND : ALL-INDIA (URBAN) AUGUST-NOVEMBER 1951

(406 urban blocks : 400 households)

vigintiles*	percentile limits of marriage duration of couples (in years)			total number of children born per couple		
	sample 1	sample 2	combined	sample 1	sample 2	combined
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	01	01	01	0.13	0.20	0.13
2	02	02	02	0.27	0.33	0.37
3	03	03	03	0.40	0.40	0.47
4	04	04	04	1.07	1.00	0.87
5	05	06	05	0.87	1.47	0.77
6	05	07	06	0.47	0.93	1.03
7	06	08	07	1.13	1.47	1.40
8	07	10	09	1.47	2.60	1.63
9	09	13	10	1.40	3.60	2.57
10	11	15	13	1.87	2.87	2.73
11	13	16	15	2.33	3.07	2.20
12	15	18	17	3.33	4.40	3.00
13	17	20	20	3.27	5.80	4.25
14	21	24	22	4.27	5.40	4.02
15	24	25	25	5.27	4.80	4.97
16	27	28	27	4.33	5.13	4.03
17	31	32	31	3.47	5.87	4.00
18	35	35	35	4.87	6.33	5.80
19	38	42	40	5.27	4.47	5.30
20	54	76	76	5.13	4.93	4.67

*See footnote to Table 5.

TABLE 7. TOTAL NUMBER OF CHILDREN BORN PER COUPLE BY DECILES OF PER CAPITA MONTHLY HOUSEHOLD EXPENDITURE AND MARRIAGE DURATION
ALL-INDIA (RURAL) : NSS 3RD ROUND
AUGUST-NOVEMBER 1951

(909 villages : 1,800 households)

per capita households expenditure (in Rs.) (deciles)	sample	marriages duration (years) (deciles)									
		1	2	3	4	5	6	7	8	9	10
(0.1)	(0.2)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	1	0.07	1.00	1.40	2.63	3.58	3.18	4.59	5.85	5.20	5.67
	2	0.17	0.76	0.73	2.00	2.44	3.56	4.39	3.77	5.28	5.50
	combined	0.09	0.53	1.36	2.17	3.08	3.40	4.44	4.38	5.30	5.59
2	1	0.06	0.54	0.50	1.63	2.00	4.43	4.75	4.50	5.64	5.60
	2	—	0.11	1.05	2.00	2.63	3.27	5.50	4.43	5.27	6.08
	combined	0.06	0.51	0.76	1.83	3.24	3.86	4.87	4.82	6.00	5.62
3	1	0.11	—	0.85	1.06	3.27	2.83	4.11	6.19	6.77	4.60
	2	0.11	0.22	1.17	1.74	2.09	2.46	4.41	5.33	5.17	5.00
	combined	0.06	0.06	1.00	1.33	1.97	3.11	4.67	5.54	4.93	4.67
4	1	0.07	0.43	1.11	1.53	2.44	4.23	5.17	5.88	5.53	6.40
	2	0.33	1.40	0.73	1.13	2.45	3.50	2.64	5.57	5.32	5.30
	combined	0.15	0.67	0.87	1.67	2.43	3.05	3.07	5.82	6.00	6.08
5	1	—	0.30	0.36	1.45	2.79	4.00	4.78	4.07	6.18	4.87
	2	0.05	0.20	0.92	1.26	3.00	3.47	4.04	4.55	5.57	4.67
	combined	0.12	0.30	0.62	1.23	3.06	3.43	4.37	4.18	5.53	4.71
6	1	0.21	0.67	0.67	1.16	1.87	3.26	4.30	4.29	4.40	5.94
	2	0.11	0.30	0.25	1.69	2.23	2.98	3.27	4.33	4.50	4.52
	combined	0.14	0.16	0.57	1.54	2.31	2.70	3.69	4.66	4.91	5.67
7	1	0.33	0.25	1.30	1.77	3.33	1.75	2.72	5.00	4.62	5.19
	2	—	0.18	1.13	1.00	2.14	2.64	4.08	4.69	5.08	6.05
	combined	0.18	0.32	1.15	1.37	2.58	3.48	4.56	3.48	4.82	5.43
8	1	0.13	1.33	1.38	1.17	2.05	4.33	3.45	4.58	5.46	9.14
	2	0.10	0.44	0.80	2.00	1.67	4.39	2.40	3.86	5.40	4.00
	combined	0.11	0.91	1.09	1.21	2.20	3.67	4.22	5.52	4.69	5.55
9	1	0.15	0.75	1.32	1.25	1.00	3.30	2.95	5.62	5.33	6.40
		0.10	0.58	1.40	0.93	1.93	3.38	3.87	4.74	4.17	4.67
	combined	0.08	0.55	1.37	0.90	1.69	3.26	3.42	4.88	5.39	5.70
10	1	—	0.33	0.94	1.26	1.73	3.10	3.19	5.38	5.70	5.75
	2	—	0.05	0.40	2.03	2.25	3.55	2.03	4.09	5.16	3.50
	combined	—	0.24	1.58	1.46	1.97	3.26	3.02	3.81	5.44	4.05

TABLE 8. TOTAL NUMBER OF CHILDREN BORN PER COUPLE BY DECILES OF PER CAPITA MONTHLY HOUSEHOLD EXPENDITURE AND MARRIAGE DURATION ALL-INDIA (URBAN) : NSS 3RD ROUND AUGUST-NOVEMBER 1951

(400 urban blocks : 400 households)

per capita households expenditure (in Rs.) (deciles)	sample	marriage duration (years) (deciles)									
		1	2	3	4	5	6	7	8	9	10
(0.1)	(0.2)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1	1	—	—	—	1.50	1.00	3.67	5.14	8.50	4.00	4.50
	2	0.00	1.20	0.00	2.00	5.00	4.00	4.00	7.50	—	3.33
	combined	0.00	1.00	0.00	1.00	3.00	4.29	4.78	7.50	4.00	3.91
2	1	0.33	0.75	1.00	2.00	4.00	3.25	2.00	4.00	8.50	5.40
	2	0.00	1.33	2.00	1.67	2.33	4.33	6.20	5.67	3.00	5.00
	combined	0.25	1.00	1.33	0.80	3.00	3.86	4.88	5.25	6.00	5.40
3	1	1.00	0.25	1.33	—	2.00	2.33	4.67	6.00	3.50	10.00
	2	0.00	0.33	—	2.00	4.50	4.00	4.83	2.75	8.00	8.75
	combined	0.33	0.25	2.00	2.00	2.38	4.00	4.67	4.00	4.40	9.00
4	1	—	3.00	1.33	0.50	1.50	—	3.17	—	1.50	2.00
	2	0.20	1.00	1.40	5.00	2.00	1.00	4.50	3.00	5.80	5.00
	combined	0.14	1.44	0.83	1.00	2.14	2.50	2.67	3.75	3.89	3.50
5	1	—	—	1.00	—	1.25	2.50	5.25	2.57	5.50	6.33
	2	0.67	1.00	0.67	1.75	2.50	6.00	3.40	5.67	1000	3.67
	combined	0.50	—	0.80	1.00	2.00	3.17	5.17	3.50	7.00	4.40
6	1	1.00	—	2.00	2.00	3.00	5.00	2.75	3.00	7.00	3.67
	2	0.20	1.00	1.67	2.25	3.25	—	13.00	—	6.38	8.00
	combined	0.40	—	1.67	3.00	2.50	3.12	4.00	4.67	7.62	3.88
7	1	0.50	—	—	—	1.75	2.25	3.00	8.00	6.33	3.50
	2	0.50	0.00	1.00	4.00	5.50	5.00	2.00	8.67	6.80	—
	combined	0.62	0.00	0.67	1.14	4.67	2.00	4.00	6.00	6.62	3.50
8	1	—	1.00	2.00	1.33	1.50	2.83	2.50	2.80	4.00	6.00
	2	0.33	0.00	1.00	2.60	2.25	6.00	7.20	4.33	4.75	4.75
	combined	0.20	0.00	1.25	2.00	1.00	2.91	4.25	4.55	4.86	5.44
9	1	0.20	0.67	0.50	—	—	6.00	—	7.50	2.50	—
	2	0.00	0.50	1.00	1.50	3.00	4.00	—	4.67	5.80	4.25
	combined	0.10	0.64	0.38	0.60	2.67	4.00	6.00	3.67	4.67	3.38
10	1	—	—	—	0.75	1.25	2.00	2.00	7.33	4.67	6.50
	2	—	0.16	2.33	3.00	3.00	1.40	3.00	4.00	5.00	4.50
	combined	0.00	0.11	1.40	1.00	2.75	2.17	3.00	5.29	4.80	4.33

Appendix 1

PROBLEM OF CONSTANT BIAS

One view is that the problem of a constant bias is not a sampling problem *per se*. However, we shall take it that one of the essential objectives of any enquiry, whether on a sample or a complete enumeration basis, is to obtain estimates with ascertainable magnitudes of different types of errors and biases.

It appears that the IPNS is generally understood very narrowly, relating only to the same level of enumerators. The technique has, of course, a much wider definition. For example, the sub-sample units may be so arranged that some are investigated by the regular primary enumerators and some others (not necessarily equal in number or size) by enumerators at a higher level, say the supervisory staff: the difference in the two sets of estimates will give an idea of the average constant bias, if any, of the primary enumerators. Similarly, the regular questionnaire may be covered by some sub-samples and a detailed questionnaire, with adequate probes, by some others (again not necessarily of the same number or size); or the reference periods may be made to differ in the different sets of sub-samples. In a demographic sample survey on births and deaths, for example, some sub-samples may be investigated by retrospective enquiries, and some others by periodic observations; some by regular enumerators, and others by medical or para-medical staff; and so on.

The different subjects under study (enumerators, methods of data collection, reference periods), when taken up in combination in different sub-samples, would represent the different combinations possible (or desirable). It is very unlikely that there will be constant bias running through *all* these combinations. And it is in this sense, that one may agree with Mahalanobis (1960) that "if all methods of ascertainment are biased to the same extent then such bias can never be detected and may be considered metaphysical."

Appendix 2

DETECTION OF ENUMERATOR DIFFERENCES

All the examples quoted did not generally reveal any differential effects of enumerators. An additional example is given below (Mahalanobis, 1946), where enumerator differences were detected by the technique of IPNS:

TABLE 9. ANALYSIS OF VARIANCE OF AGE (IN YEARS): BENGAL LABOUR ENQUIRY, JAGADDAL, 1944[a]

(3 Blocks × 5 Enumerators)

source	d.f.	m.s.	*F*-ratio
(0)	(1)	(2)	(3)
Block	2	62.1	0.49
enumerator	4	304.8	2.39*
block × enumerator	8	78.5	0.61
between sub-samples	14	140.8	1.10
within sub-samples	510	127.7	—
total	524	128.1	—

[a] Mahalanobis (1946), with equalized number of families per cell.
* Significant at 5% level.

Appendix 3

COMPUTATION OF THE LOSS OF INFORMATION

The formula given by Mokashi (1950) for the loss information per unit cost due to IPNS is:

$$1-\sqrt{\frac{c_j+c_k}{2c_j+c_k}}$$

for randomly allocated samples, as followed in the Indian NSS, where c_j = cost per journey time, and c_k = cost per other time (enumeration, statistical work etc.), per square mile. From the NSS eighth round, on basis of the time records kept by the investigators, it was estimated that $c_j/c_k = 0.10$, from which the loss of information per unit cost was obtained as 3½ per cent.

REFERENCES

AGGARWAL, O. P. (1963): *Primer Censo Nacional Aggropecuario 1961*. Republica del Peru, Instituto Nacional de Planification, Lima (mimeographed).

COCHRAN, W. G. (1953): *Sampling Techniques*, John Wiley & Sons, Inc., New York.

DAS GUPTA, A; SOM, R. K., MAJUMDAR, M. and MITRA, S. (1955): Couple fertility. *National Sample Survey*, Report No. 7, Government of India.

DAS GUPTA, A. and SOM, R. K. (1959): Some characteristics of the economically active population, Government of India. *National Sample Survey*, Report No. 14

FEDERATION OF RHODESIA and NYASALAND, CENTRAL AFRICAN STATISTICAL OFFICE (1951): *Report on the Demographic Sample Survey of the African Population of Southern Rhodesia*, 1948.

——— (1959): *The 1953-1955 Demographic Sample Survey of the Indigenous African Population of Southern Rhodesia.*

HANSEN, M. H., HURWITZ, W. N. and BERSHAD, M. (1961): Measurement errors in censuses and surveys. *Bull. Int. Stat. Inst.*, 38, 359.

HANSEN, M. H., HURWITZ, W. N. and MADOW, W. G. (1953): *Sample Survey Methods and Theory*, John Wiley & Sons, Inc., New York.

HANSEN, M. H., HURWITZ, W. N., MARKS, E. S. and MAULDIN, W. P. (1951): Response errors in surveys. *J. Amer. Stat. Assoc.* 46, 147.

INDIAN STATISTICAL INSTITUTE (1962): Rates of birth and growth of rural population: Fifteenth round, July 1959—July 1960, *National Sample Survey*: Draft Report No. 110.

KOOP, J. C. (1960): On theoretical questions underlying the technique of replicated or interpenetrating samples. *Proc. Social Stat. Sec., A.S.A.*, 196.

LAHIRI, D. B. (1954): Technical note on some aspects of the development of the sample design. *National Sample Survey*, Report No. 5, Government of India.

MAHALANOBIS, P. C. (1938-41): *A Statistical Report on the Experimental Crop Census of* 1937; subsequently printed in the series, *On the Sample Survey of Jute in Bengal*, Indian Central Jute Committee, Calcutta, 1938-41.

——— (1946): Recent experiments in statistical sampling in the Indian Statistical Institute. *J. Roy. Stat. Soc.*, 109 (4), 325; reprinted Asia Publishing House and Indian Statistical Institute (1960a).

——— (1950): Cost and accuracy in sampling and complete enumeration. *Bull. Int. Stat. Inst.*, 32, 210.

——— (1960b): A method of fractile graphical analysis. *Econometrica*, 28(2), 325.

MAHALANOBIS, P. C. and DAS GUPTA, A. (1954): The use of sample surveys in demographic studies in India. *World Population Conference, Rome*, 1954, E/Conf. 13/131.

MOKASHI, V. K. (1950): A note on interpenetrating samples. *J. Ind. Soc. Agr. Stat.* 2, 189.

MURTHY, M. N. and NANJAMMA, N. S. (1959): Almost unbiassed ratio estimates based on interpenetrating sub-sample estimates. *Sankhyā*, 21, 299.

POTI, S. J., RAMAN, M. V., BISWAS, S. and CHAKRAVORTY, B. (1959): A pilot health survey in West Bengal: 1955. *Sankhyā*, 21, 141.

RAMACHANDRAN, K. V. (1958): On the Studentized smallest Chi-square. *J.A.S.A.*, 53(284), 868.

SATTERTHWAITE, F. E. (1946): An approximate distribution of estimates of variance components. *Biometrics*, 2, 110.

SOM, R. K. and MUKHERJEE, H. L. (1962): Analysis of variance of demographic variables. *Sankhyā*, Series B, 24, 13.

SOM, R. K., DE, A. K., DAS, N. C., PILLAI, B. T., MUKHERJEE, H. and SARMA, S. M. U. (1961): Preliminary estimate of birth and death rates and of the rate growth of population, Fourteenth round, July 1958-1959, Government of India. *National Sample Survey*, Report No. 48

SOM, R. K., DAS, N. C. and DE, A. K. (1961): Report on morbidity, Government of India. *National Sample Survey*, Report No. 49

SUKHATME, P. V. (1953): *Sampling Theory of Survey with Applications*. The Iowa State College Press, Ames and the Indian Society of Agricultural Statistics, New Delhi.

SUKHATME, P. V. and SETH, G. R. (1952): Non-sampling errors in surveys. *J. Ind. Soc. Agr. Stat.*, 4, 5.

UNITED NATIONS (1960): *A short manual on sampling, Vol. I, Elements of sample survey theory*. Studies in Methods, Series F, No. 9.

ZARKOVICH, S. S. (1963): *Sampling Methods and Censuses, 2, Quality of statistical data.* Food and Agricultural Organization of the United Nations.

Paper received: May, 1965.