SELF-WELIGHTING DESIGN AT TABULATION STAGE*

By M. N. MURTHY
Indian Statistical Institute
and
V. K. SETHI
Institute of Social Sciences, Agra

SUMMARY. When a non-self-weighting design is used at the field stage in a large scale survey, the work at the tabulation stage becomes time consuming and costly due to the large number of multipliers (inflation factors) involved in obtaining the estimates. In this paper some techniques of making the design self-weighting, at least partially, at the tabulation stage have been considered. One of these techniques, which is likely to be the most efficient in practice, is the selection of a sub-sample of the units in the sample with probability proportional to the multipliers systematically after some suitable arrangement of the units such that almost all the units in the sample are included in the sub-sample at least once.

1. Introduction

One of the difficulties in making the sample design self-weighting at the field stage is that the work-load in the penultimate stage units may become unequal which is not desirable from the point of view of administrative considerations in large scale surveys. The methods which make the sample design self-weighting at the field stage ensuring at the same time equal work-load within penultimate stage units impose rather severe restrictions on the allocation and selection procedures. Because of these considerations it may neither be desirable nor feasible to adopt a self-weighting design at the field stage in certain situations.

Suppose n ultimate stage units are selected from a population according to some specified non-weighting design. Let y_i be the value of the characteristic of the i-th selected sampling unit and let w_i be the corresponding multiplier (inflation factor or weight) to be used in getting an unbiased estimate of the population total. Then the estimate of the population total Y is given by

$$\hat{Y} = \sum_{i=1}^{n} w_i y_i. \qquad \ldots \quad (1.1)$$

The problem is to find a technique by which the set of multipliers can be replaced by a smaller number of multipliers such that the estimate remains unbiased and the increase in variance is the least or if the estimate becomes biased the bias is negligible. This problem has been considered by the authors in an earlier paper (1961).

* Originally issued in 1959 as a mimeographed publication by the Department of National Sample Survey, Indian Statistical Institute.

2. Rounding-off techniques

The following five procedures of making the design self-weighting at the tabulation stage are considered here :

(i) rounding-off of the multipliers to the nearest multiple of a convenient number such as hundred, thousand or ten thousand;

(ii) substituting each multiplier by the mean of the multipliers;

(iii) rounding-off of multipliers to an optimum set of rounded-off multipliers;

(iv) sub-sampling with probability proportional to the multipliers with replacement; and

(v) sub-sampling with probability proportional to the multipliers systematically.

Procedures (i) and (ii) give biased estimates with possible decrease in the variance of the estimate under certain circumstances, while procedures (iii), (iv) and (v) give unbiased estimates with some increase in the variance. It may be mentioned that in some surveys procedures (i) and (v) are being adopted to expedite the tabulations.

Procedure (i). Suppose we have a series of four digit multipliers. The usual procedure of decreasing the number of multipliers is to round off the multipliers to the nearest thousand. Thus the n four digit multipliers are replaced by at most ten rounded-off multipliers. The estimate obtained by using this method will obviously be biased. The magnitude of bias depends on the multipliers as well as on the values of the characteristics. In practice, it is difficult to get an idea of the sign and magnitude of the bias unless one works out the biased estimate as well as the estimate using the actualm ultipliers for at least some selected characteristics.

Procedure (ii). If this method is adopted, the estimate is given by the product of the mean of the multipliers and the sum of the values of the characteristic. This estimate is also biased, but the bias will be negligible if in the sample, the covariance between the multiplier and the value of the selected unit is very small. For instance, in a particular per capita expenditure class, it is felt, the expenditure on cereals, food and total expenditure of a household might not depend much on its multiplier. In other words, the expenditure on cereals, food and total expenditure of a household, belonging to a particular per capita expenditure class are expected to depend more on the household size, the geographical, the climatic and the economic conditions of the locality than on whether the household is in a large village or a small village. If only the estimates of the expenditure on cereals and such items are required, the procedure would be to classify the households according to the per capita expenditure class and then find the biased estimates separately for all classes. The sum of these biased estimates will be the estimate of the population total and the bias can be xpeeted to be small. It may be noted that in this case the number of multipliers will be equal to the number of per capita expenditure classes.

In this connection it should also be noted that this method is not to be used indiscriminately as there may be characteristics which are related to the multipliers. This method can be used only if we are fairly certain that the value of the characteristic in which we are interested is uncorrelated with the multiplier.

Procedure (iii). A general solution would be to round off each of the multipliers to a certain number of weights, which may be called rounded-off multipliers with such probabilities that the expected value is the original multiplier. As these weights are at our choice, we can choose them such that the increase in variance is minimized. Further a prespecified level of increase in the variance at the tabulation stage can be achieved by taking a sufficient number of rounded-off multipliers in the optimum fashion.

As the optimum solution for a specified number of rounded-off multipliers depends on the value of the characteristics in question, it is not feasible in practice to get the optimum solution for each characteristic, though some method which would give us a solution near about the optimum can be devised. Another disadvantage is that even the determination of the approximate optimum rounded-off multipliers becomes increasingly difficult as the number of such multipliers is sought to be increased. This procedure is considered in detail by the authors in an earlier paper (1961).

Procedure (iv). This method consists in taking a sub-sample of n' units from the field sample of size n with probability proportional to the multipliers with replacement. The sum of the values in the sub-sample multiplied by a constant (the ratio of the sum of n multipliers to n') gives an unbiased estimate of $\sum_{i=1}^{n} w_i y_i$. The increase in variance at the tabulation stage is given by

$$E_f \frac{1}{n'} \left[\left(\sum_{i=1}^{n} w_i \right) \left(\sum_{i=1}^{n} w_i y_i^2 \right) - \left(\sum_{i=1}^{n} w_i y_i \right)^2 \right] \qquad \dots \quad (2.1)$$

where E_f is the expected value over the field sample. It can be seen that this method results in rounding off each of the multipliers to one of the set of multipliers $\left(j \frac{\sum_{i=1}^{n} w_i}{n'} \right)$, $j = 0, 1, 2, \dots, n'$, with certain probabilities such that the estimate remains unbiased and that the sum of the rounded-off multipliers in the sub-sample is equal to the sum of the actual multipliers.

If instead of sub-sampling, we round off each of the multipliers to one of the set of multipliers $\left(j \frac{\sum_{i=1}^{n} w_i}{n'} \right)$, $j = 0, 1, 2, \dots, n'$ with probabilities

$$\binom{n'}{j} \left(\frac{w_i}{\sum_{i=1}^{n} w_i} \right)^j \left(1 - \frac{w_i}{\sum_{i=1}^{n} w_i} \right)^{n'-j}, \; i = 1, 2, \dots, n,$$

then the estimate is unbiased and the increase in variance is given by

$$E_f \frac{1}{n'}\left[\left(\sum_{i=1}^{n} w_i\right)\left(\sum_{i=1}^{n} w_i y_i^2\right)-\left(\sum_{i=1}^{n} w_i^2 y_i^2\right)\right]. \qquad \dots \quad (2.2)$$

Comparing (2.1) and (2.2) we see that the expression (2.2) is greater than (2.1) since $w_i y_i$'s are positive in general. Hence sub-sampling with probability proportional to the multipliers with replacement is more efficient than the corresponding randomizing method. The sub-sample size n' can be so fixed as to almost completely avoid rejections of the sample units provided such a procedure does not lead to considerable repetitions of many of the units.

Procedure (v). In this method the units in the field sample are arranged in some suitable order, the multipliers are cumulated and a systematic sample with probability proportional to multipliers is drawn with the interval $(\sum_{i=1}^{n} w_i/n')$. As in the case of probability proportional to the multipliers with replacement, the estimate is given by the product of the sum of the values in the sub-sample and a constant which is the sampling interval. An expression for the increase in variance is difficult to find in this case. This method results in rounding off each of the multipliers (say, w_i) to

$$\left[\frac{n'w_i}{\sum_{i=1}^{n} w_i}\right]\frac{\sum_{i=1}^{n} w_i}{n'} \quad \text{or} \quad \left[\frac{n'w_i}{\sum_{i=1}^{n} w_i}+1\right]\frac{\sum_{i=1}^{n} w_i}{n'}$$

with certain probabilities such that the estimate remains unbiased and that the sum of the rounded-off multipliers in the sub-sample is equal to the sum of multipliers in the field sample.

A sort of 'without replacement' element is present in this systematic selection with probability proportional to the multipliers. For, when the units are arranged at random and the sizes are equal, this procedure amounts to simple random sampling without replacement, while in the case the probability proportional to multipliers with replacement procedure amounts to simple random sampling with replacement. Further, drawing of a sub-sample systematically with probability proportional to the multipliers is likely to take less time than in the corresponding with replacement sampling. The estimate in case of pps systematic method can be improved upon by arranging the units in some suitable order and by devising appropriate balancing procedures.

The randomizing method corresponding to pps systematic selection is more efficient than the corresponding randomizing method of pps with replacement selection. For, in the former each multiplier is rounded off to the nearest two rounded-off multipliers on either side of it and it can be easily shown that in such a case the variance will be the least (cf. Appendix I). We have already shown that pps with replacement sub-sampling is more efficient than the corresponding randomizing procedure. This is probably due to the fact that in the case of pps with replacement the sum of the rounded-off multipliers in the sub-sample is equal to the sum of the multipliers in the field sample, whereas this is not achieved in the corresponding randomizing procedure.

In pps systematic sub-sampling both the desirable conditions mentioned above, namely, the sum of the rounded-off multipliers in the sub-sample being equal to the sum of the multipliers in the field sample and each multiplier being rounded off to one of the two nearest rounded-off multipliers on either side of it such that the expected value is w_i are satisfied. Hence it can well be expected that the pps systematic selection will be more efficient than the pps with replacement selection. It may be mentioned that in pps systematic sampling the expected number of repetitions of any unit would be proportional to its size.

3. An empirical study

As it is not possible, in general, to compare the efficiencies of the procedures suggested in Section 2 to reduce the cost at the tabulation stage, an empirical study was conducted to assess their merits and demerits. For this purpose the data on consumer expenditure statistics collected in a large scale sample survey in Uttar Pradesh were used. The object of this study was to compare the efficiencies of the procedures (i), (ii), (iv) and (v) as well as their practicability in large scale operations, in estimating (a) expenditure on cereals, (b) expenditure on food and (c) total expenditure by per capita expenditure classes.

The design of the survey was a stratified three stage one with tehsils as first stage units, villages as second stage units and households as the third stage units. From each stratum two tehsils were selected with probability proportional to population with replacement. From each selected tehsil two villages were selected with probability proportional to population with replacement and from each selected village five households were selected systematically with a random start for the consumer expenditure enquiry. It is to be noted that for each stratum total we get two independent estimates, one from each of the two tehsils selected from that stratum. The sample households belonging to the tehsils selected first will be considered as belonging to the field sample 1 and the other sample households will constitute the field sample 2.

The sample households belonging to each of the field samples were grouped into fourteen classes on the basis of their per capita total expenditure. In each per capita expenditure class the households were arranged according to the village, tehsil and stratum to which they belonged.

For procedures (iv) and (v), from each per capita expenditure class a sample of size equal to the number of sample households in that class was selected to estimate the total expenditure on cereals, food and total expenditure. This sample size was taken in each class, since in that case the efficiencies of procedures (iv) and (v) become comparable with those of procedures (i) and (ii). The variances of the estimates obtained by using procedures (iv) and (v) were calculated assuming the field sample to be the population. These variances can be considered as the estimates of the increase in the variances of the estimates of the population totals due to sub-sampling. The bias of the estimates based on procedures (i) and (ii) were also calculated.

Let w_{ij} and y_{ij} be the multiplier and the variate value for the j-th sample household of the i-th per capita expenditure class. The object is to estimate $y_i = \sum_{j=1}^{n_i} w_{ij}y_{ij}$ where n_i is the number of households in the i-th class. The variance of the estimate based on a sample taken with probability proportional to multipliers with replacement is given by

$$V_{pps}(\hat{y}_i) = \frac{1}{n_i}\left[\left(\sum_{j=1}^{n_i} w_{ij}\right)\left(\sum_{j=1}^{n_i} w_{ij}y_{ij}^2\right)-\left(\sum_{j=1}^{n_i} w_{ij}y_{ij}\right)^2\right]. \quad \ldots \quad (3.1)$$

The variance of the estimate based on the pps systematic (size being the multiplier) sampling procedure is obtained by using the method illustrated in Appendix 2. The estimates in case of procedures (i) and (ii) are respectively

$$\hat{y}_i = \sum_{j=1}^{n_i} r_{ij}y_{ij} \quad \ldots \quad (3.2)$$

and

$$\hat{y}_i = \frac{1}{n_i}\left(\sum_{j=1}^{n_i} w_{ij}\right)\left(\sum_{j=1}^{n_i} y_{ij}\right). \quad \ldots \quad (3.3)$$

where r_{ij} is the nearest multiple of 10000 of the multiplier w_{ij}. This estimate summed over all the classes will give us an estimate of the total for the population as a whole. It is expected that for procedure (ii) the bias in the estimate of the population totals considered here is likely to be small if the sample households are stratified according to per capita expenditure class at the tabulation stage. The differences between these estimates and $\sum_{j=1}^{n_i} w_{ij}y_{ij}$ $(i=1,2, \ldots, 14)$ give the biases.

The results of the empirical study are given in Table 1. From Columns (4) and (5) of this table we find that the pps systematic estimate is much better than the pps with replacement estimate in all the cases. It is possible to improve upon the pps systematic estimate by suitable arrangement of the sample units before sub-sampling. Of course in the case of pps systematic sampling the variance of the estimate cannot be estimated unbiasedly from one sample. However this difficulty can be overcome by taking two or more sub-samples with independent random starts.

The comparison of procedures (i) and (ii) with procedure (iv) shows that the biased procedures may be preferred. Procedures (i) and (ii) should be used with considerable caution as they give rise to biased estimates and it is difficult to assess the magnitude of bias in practice. Procedure (ii) will be a good method, as has been pointed out earlier, only if the correlation coefficient between the multipliers and the variate values is very nearly equal to zero. The pps systematic estimate compares favourably with this biased estimate. Since the pps systematic method has the added advantage of being unbiased, it may be preferred.

Acknowledgement

The authors are grateful to Professor D. B. Lahiri for suggesting the problem.

TABLE 1. COMPARISON OF PERCENTAGE BIASES OF PROCEDURES (i) AND (ii) WITH PERCENTAGE STANDARD ERRORS OF PROCEDURES (iv) AND (v)

expenditure class	sample 1: sample size	sample 1: percentage bias (i)	sample 1: percentage bias (ii)	sample 1: percentage s.e. (iv)	sample 1: percentage s.e. (v)	sample 2: sample size	sample 2: percentage bias (i)	sample 2: percentage bias (ii)	sample 2: percentage s.e. (iv)	sample 2: percentage s.e. (v)
(0)	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
				characteristic : expenditure on cereals						
1	4	− 2.71	+27.41	43.38	15.47	9	+ 9.00	+ 2.47	15.00	7.82
2	18	+10.22	− 8.62	15.44	3.15	19	+ 0.15	− 1.67	16.24	4.70
3	55	+ 1.86	− 2.84	7.16	0.84	43	− 5.63	+ 0.70	8.78	4.86
4	31	− 0.60	+ 1.15	8.74	1.21	41	+ 7.28	− 2.60	30.67	5.62
5	38	− 2.24	− 3.07	7.97	1.00	45	− 0.78	− 1.08	7.69	3.12
6	58	+ 3.79	− 1.89	7.60	1.11	62	+ 2.72	− 4.10	7.20	2.52
7	33	− 1.88	− 4.70	8.01	1.88	38	− 3.55	− 5.48	11.60	4.64
8	33	− 2.09	− 9.73	10.40	1.63	24	+ 3.00	− 6.53	19.48	5.68
9	23	− 1.47	+ 0.49	11.81	2.83	21	− 9.07	− 0.34	15.77	13.06
10	10	+12.48	− 5.61	17.44	2.52	22	+ 5.15	− 0.32	15.44	7.26
11	24	+ 5.70	+ 2.74	13.08	1.72	16	− 4.55	+ 4.31	22.20	13.72
12	10	− 2.11	−20.90	24.28	3.17	5	+17.55	− 4.39	34.86	12.50
13	8	+ 1.10	+ 1.41	16.64	2.43	5	+14.87	− 2.91	24.29	11.00
14	4	+ 6.45	− 4.95	20.34	3.53	2	−26.62	+ 1.20	33.38	7.96
all	358	+ 1.47	− 3.36	3.10	1.03	341	+ 0.69	− 2.36	5.70	1.04
				characteristic : expenditure on food						
1	4	− 0.68	+33.70	47.17	18.07	9	+ 7.85	+ 1.28	18.38	9.05
2	18	+11.44	− 5.25	12.87	1.04	10	+ 3.61	− 0.38	10.84	2.04
3	55	+ 2.49	− 0.46	5.76	0.92	43	− 4.58	+ 1.96	8.61	3.82
4	31	− 1.32	+ 0.18	7.25	0.95	41	+ 5.67	− 0.31	24.46	4.52
5	38	− 3.16	− 0.05	6.25	0.91	45	− 0.50	+ 0.08	6.88	4.64
6	58	+ 5.47	+ 1.97	6.36	0.93	52	+ 3.44	− 2.24	6.38	2.74
7	33	− 0.10	− 3.16	7.52	1.60	38	− 0.95	− 0.37	8.99	3.88
8	33	− 1.25	− 0.87	9.08	1.57	24	+ 5.39	− 2.00	16.84	6.04
9	23	− 0.13	+11.80	10.95	3.63	21	− 7.72	+ 3.86	16.88	6.06
10	19	+11.39	− 4.37	12.24	2.41	22	+ 7.87	+ 2.10	11.93	4.68
11	24	+ 3.13	− 2.08	10.71	2.51	15	− 4.84	+ 5.12	23.17	16.04
12	10	− 2.65	−37.53	30.04	3.48	5	+13.56	− 3.44	21.92	6.98
13	8	− 3.26	− 5.61	10.09	3.40	5	+16.27	− 4.19	27.07	10.60
14	4	+ 3.18	+ 1.45	18.23	2.50	2	+26.91	+ 1.00	24.45	6.22
all	358	+ 1.51	− 2.06	3.07	1.16	341	+ 1.63	− 0.00	4.72	1.66
				characteristic: total expenditure						
1	4	− 0.94	+33.33	46.07	17.00	9	+ 8.05	2.03	20.47	10.62
2	18	+11.22	− 5.04	12.31	1.85	10	+ 3.69	−0.42	10.84	2.42
3	55	+ 2.10	− 0.29	5.40	0.85	43	− 4.27	−1.68	9.62	3.00
4	31	− 1.20	+ 0.42	7.15	0.85	41	+ 5.22	−0.28	23.31	2.10
5	38	− 3.44	+ 2.38	6.68	0.97	45	− 1.21	1.19	7.05	5.22
6	58	+ 5.42	+ 2.40	6.55	1.05	62	+ 4.35	−3.68	8.60	2.50
7	33	− 0.06	− 3.51	7.50	1.87	38	− 0.82	−1.88	7.75	3.02
8	33	− 1.19	− 8.72	9.03	1.84	24	+ 4.50	−1.78	10.65	6.24
9	23	− 0.57	+ 3.30	11.91	2.44	21	− 8.08	12.44	10.22	7.52
10	10	+11.88	− 2.52	11.23	2.08	22	+ 7.37	1.69	11.29	6.14
11	24	+ 2.75	+ 2.84	11.20	3.18	15	− 4.45	2.62	21.70	11.10
12	10	− 2.66	−30.33	25.40	3.10	5	+16.78	−5.14	20.41	10.30
13	8	− 1.65	− 1.34	20.46	4.30	5	+21.98	−6.78	48.75	23.64
14	4	− 3.37	+13.17	30.86	10.66	2	+26.58	1.43	35.22	8.80
all	358	+ 1.04	− 1.63	3.61	1.62	341	+ 2.30	+ 0.13	4.89	2.34

Sample size: number of households in the sample;
(i) Rounding off the multipliers to the nearest multiples of ten thousand;
(ii) Substitution by mean of the multipliers;
s.e.: standard error;
(iv) pps with replacement
(v) pps systematic.

Appendix 1

THEOREM: *Let S be a set of random variables each taking values belonging to a given set T and having X as the expected value. Then of all the members of S, that random variable x taking only the two values of the set T which are on either side of X and are nearest to it has the least variance.*

Proof: Suppose x_j and x_{j+1} are the nearest two values belonging to T on either side of X. Let x take values x_j and x_{j+1} with probabilities p and $(1-p)$ respectively. Since x belongs to the set S, $E(x) = X$ and hence $p = (x_{j+1}-X)/(x_{j+1}-x_j)$. Let $x_{j+1}-x_j = a$ and $X = x_j+b$. Then the variance of x is

$$V(x) = (x_{j+1}-X)(X-x_j) = b(a-b).$$

Let y denote any member of S other than x and let y take the values (x_i) with probabilities (p_i), $(i = 1, 2, \ldots, n)$. The variance of y is given by

$$V(y) = \sum_{i=1}^{n} (x_i - X)^2 p_i$$

where $\sum_{i=1}^{n} p_i = 1$ and $\sum_{i=1}^{n} x_i p_i = X$. Substituting x_j+b for X in the variance of y we get

$$V(y) = \sum_{i=1}^{n} (x_i-x_j)^2 p_i - b^2.$$

Variance of x can be written as

$$V(x) = a\sum_{i=1}^{n} (x_i-x_j) p_i - b^2.$$

Hence we find that $V(y) > V(x)$, if $\sum_{i=1}^{n} (x_i-x_j)^2 p_i > a\sum_{i=1}^{n} (x_i-x_j) p_i$. A sufficient condition for this is that

$$(x_i-x_j)^2 > a(x_i-x_j)$$

that is,

$$(x_i-x_j)(x_i-x_{j+1}) > 0$$

which is true for there is no x_i lying between x_j and x_{j+1}. Hence $V(y) > V(x)$. As this is true for all y, we see that x has the least variance.

Appendix 2

A procedure is given below for calculating the variance of the estimate based on a pps systematic sub-sample. This method is based on the fact that the number of different pps systematic samples will be n or $n+1$, if n, the number of units in the population, is less than the sampling interval. The number of different samples will be less than n, if the sampling interval is less than n or if some of the sizes are equal to or multiples of the sampling interval. In this method we directly get the different possible samples with their frequencies and making use of these the variance of the estimate can be found.

Let w_i $(i = 1, 2, \ldots, n)$ be the multiplier for the i-th unit in the sample. The sampling interval for getting a pps systematic sub-sample of size n' is $\sum_{i=2}^{n} w_i/n'$ $(= I$, say). As has been pointed out earlier the i-th unit will be repeated k_i $(=[w_i/I])$ or k_i+1 times in the sub-sample $(i = 1, 2, \ldots, n)$. Let r_i be the remainder obtained by dividing the cumulated total of the multipliers against the i-th unit by I. If $r_i > r_{i-1}$, the i-th unit will be repeated k_i+1 times provided the random start R lies in the interval $(r_{i-1}+1, r_i)$ and k_i times otherwise. This situation will be denoted by $((r_{i-1}+1, r_i)\ k_i+1; k_i)$

If $r_i < r_{i-1}$, then we get $((r_i+1,\ r_{i-1})\ k_i,\ k_i+1)$. If $r_i = r_{i-1}$, then $((1,\ I)\ k_i)$. In practice the calculation of the r_i's will be fairly simple. First r_1 is obtained by dividing w_1 by I. To this r_1, w_2 is added and r_2 is the remainder obtained by dividing (r_1+w_2) by I and so on. An illustration of the application of this method to sub-sampling of households from sample households in a per capita expenditure class with probability proportional to multipliers systematically is given in Tables A1 and A2.

TABLE A.1. MULTIPLIERS AND VALUES OF EXPENDITURE ON FOOD FOR SAMPLE HOUSEHOLDS WITH NUMBER OF REPETITIONS IN A PPS SYSTEMATIC SUB-SAMPLE OF SIZE 9 FOR DIFFERENT POSSIBLE RANDOM STARTS

sl. no.	multipliers	expenditure on food	number of repetitions and range of interval
(0)	(1)	(2)	(3)
1	45696	21.93	(1, 8132) 2; 1
2	45696	19.62	(8133, 16264) 2; 1
3	32590	16.62	(11291, 16264) 0; 1
4	32590	7.19	(6317, 11290) 0; 1
5	30068	13.50	(6317, 36384) 1; 0
6	30068	50.23	(28889, 36384) 0; 1
7	33080	13.31	(28889, 29404) 2; 1
8	43726	18.79	(3003, 29404) 1; 2
9	34560	13.68	(3003, 33502) 1; 0

Here the sub-sample size is taken to be 9 and hence the sampling interval is $I = \sum_{i=1}^{9} w_i/9 = 37564$. From Table A.1 it follows that if the random start lies in the interval (1, 3002) we get the sub-sample $(u_1\ u_1\ u_2\ u_2\ u_4\ u_6\ u_7\ u_8\ u_8)$ where u_i is the i-th sample household. Hence this sample occurs with a frequency of 3002. Similarly proceeding we get the different possible samples (s) with their frequencies (f_s) shown in Table A.2. The sample totals given in Column (4) of Table A.2 multiplied by I will give us an unbiased estimate y_s of $\sum_{i=1}^{9} w_i\ y_i$. The mean of all the sub-sample estimates is given by

$$\frac{\sum_{s=1}^{10} f_s y_s}{\sum_{s=1}^{10} f_s} = 6486 \times 10^3$$

which is equal to the weighted estimate verifying the fact that the estimate from pps systematic sub-sample is unbiased. The variance of the estimate is given by

$$V(y_s) = \frac{\sum_{s=1}^{10} f_s y_s^2}{\sum_{s=1}^{10} f_s} - y^2 = 344320(10)^4$$

TABLE A.2. POSSIBLE SUB-SAMPLES AND THEIR FREQUENCIES WITH SUB-SAMPLE TOTALS FOR EXPENDITURE ON FOOD

sub-samples	interval range	frequency f_s	units in sub-sample with repetitions, if any	expenditure on food (sub-sample total)
(0)	(1)	(2)	(3)	(4)
1	(1, 3002)	3002	1 1 2 3 4 6 7 8 8	188.41
2	(3002, 6316)	3314	1 1 2 3 4 6 7 8 9	183.30
3	(6173, 8132)	1816	1 1 2 3 5 6 7 8 9	169.61
4	(8133, 11290)	3158	1 2 2 3 5 6 7 8 9	187.30
5	(11291, 16264)	4974	1 2 2 4 5 6 7 8 9	177.87
6	(16265, 28888)	12624	1 2 3 4 5 6 7 8 9	174.67
7	(28889, 29404)	516	1 2 3 4 5 7 7 8 9	137.95
8	(29405, 36384)	6980	1 2 3 4 5 7 8 8 9	143.43
9	(36385, 37562)	1178	1 2 3 4 6 7 8 8 9	180.16
10	(37563, 37564)	2	1 2 3 4 6 7 8 8	166.48

REFERENCE

MURTHY, M. N. and SETHI, V. K. (1961): Randomized rounded-off multipliers in sampling theory. *J. Amer. Stat. Ass.*, 56, 328-334.

Paper received : December, 1962.