SELF- WEIGHTING DESIGN FOR A-STRATIFIED SAMPLING

BY
DR. MOHAMED ZAID
Zagazing University, Benha Branch, Faculty of Commerce.

ABSTRACT. The aim of this paper is to introduce a stratified two - stage self-weighting desigdn which maxamises the number of selected ultimate sampling units when the number of primary sampling units is small. A method of allocation of a given number of primary sample units strata, which maximizes the sample size of ultimate units is discussed in this paper.

1. Introduction

Consider a stratified two - stage sampling design where the self-weighting is achieved by selecting the primary sampling units (psu's) by simple random sampling without replacement and the second - stage sampling units (ssu's) by linear systematically sampling with pre-determined intervals. The concerned inquiry is relevant for only a few ssu's belonging to a particular class. An unbiased estimate of the total for an item in the universe is required to he obtained.

It is known that no remarkable gain in sampling efficiency is achieved if the average number of ssu's selected per sample psu in s-th stratum kh,) is more than 6 for a given number of sample pus's

, and a given $= \frac{\text{ratio*}}{A,,}$

Let us assume that the average number of ssu's per psu in s-th stratum is much less thanb, say, it is 2. Then naturally we require all

---

* A_i, and $A.$ are the terms contained in the two — stage sampling

variance A $= \frac{A,}{17} + \frac{A}{17.\,17,}$ A. AS h increases, the ratio $A,$

decreases to attain the stage of no remarkable gain.

these ssu's to be surveyed in a sample psu In other words, it becomes our objective to maximize the expected total number of ssu's to corresponding a given number of psu's .This situation occurs in practice if the survey is confined to the households, engaged in infrequent activities such as mining or construction, constituting the ultimate sampling units.

In a self - weighting design, the weighting multiplier can be calculated in tow Ni a ys. Firstly, we can take $\frac{}{17}$ as the multiplier for all - strata level where H is an estimate of a total number of concerned (ssu's) over all the strata obtained from an earlier census or survey and h , the expected total number of ssu's planned to be selected from them. Alternativel), we can take the greatest stratum - psu tnultiplier as the multiplier for all- strata. In the concerned situation, as the over -all sampling fraction has to be made large enough to get reliable estimates, the multiplier obtained from the first method ma) become ery small and hence the interval for sampling ssu's *be* Jess than one unit for at least some pus's creating a problem in keeping the design self- weighting.

2. The self - weighting design

The unbiased estimate Y of the population total for item Y of ssu's concerned is given by

$$Y = \quad -t\, n, \tag{2-1}$$

Where $n,psu's$ are selected b) simple random sampling without replacement (srss, or) from N, psu's of s- th stratum, s=1,2,...K . after adopting proportional allocation of sample psu's (rounded up to the nearest positis e integer) to strata under the constraints,

$$\hat{N} = \quad 111 = 17 \text{ and } 17 \quad K, \quad \text{and} \quad h \quad ssu's,$$

and tic, ssu's are selected by linear s)stematicall) sampling from total H concerned ssu's with interval /,, (rounded up to the nearest

nearest decimal place) in si-th sample psu $n,)$ .The subscript j stands for the ssu.

A sampling design is called self-weighting, at al strata level, if for estimating the total of an item there is a single t :•;iinion weight (or multiplierM, say) for each of the ssu's in the sample- The problem of making the design self-weighting at field stage w discussed by Hansen, Hurwitz and Madow (1953). and Lahiri (1954; Al is usually determined from where H, the total number of sst.'s in all the strata, is estimated from an earlier census or survey aim it is the expected total number of ssu's to be selected from H.

Another in this paper the multiplier is obtained by a new meth of allocation in the following manner. The vector v„ = ,)

Satisfying En„,=n is defined as an optimum integral allocation for a given vector V = $(N,,N_{„},...,N_{A})$ if the maximum stratum - level psu multiplier is minimum over. All integral allocations v s, where v =)satisfies1 = 17 and are positive integers denoting the number of psu's to be selected from s-th stratum. An allocation v_0 is optimum if max $\frac{A'}{17_{3I}}$ corresponding to it is not greater than max $=$ corresponding to anyv. In other words, max $\frac{}{17_{11}}$==

The procedure for obtaining v„ is discussed in Section 3.

After obtaining v_0, $M_{„}$= max --I is taken as the multiplier for all-strata. So, there is no need for having prior information on 11 for the pre-determination of multiplier in this case.

Then, 1 under v_o or $/_{0,,,}$ is determined from

$$= M_{II} \qquad (2,2)$$

Obviously, 1 the same for all psu's in the same stratum. All theik, ssu's in si-th psu are listed and then from them ssu's are selected using the interval ./„. Will be 1.0 in at least one stratum but will never be less than 1.0 in any stratum. Use of any multiplier less than M„ makes / less than 1.0 in at least one stratum and so the self-weighting design fails. Su, Al„ is the minimum multiplier which maximizes the number of ssu's. The expression for the unbiased estimate Ygiven by (2,1) where n_s and ./,` are to be are to be replaced by n_{ip}, and . respectively now assumes the form.

$$I = \qquad (2,3)$$

3- Procedure kit finding I .„

It is well known that, to have constant intervals, proportional allocation (usually rounded up to the nearest positive integers) is required in order to keep the design self-weighing. The purpose of achieving optimum integral allocation in a self-weighing design is to have intervals of length close to the constant value 1.0. It follows that the allocation v„ is expected to lie in the neighborhood of roportional allocation. Hence, proportional allocation has been generally taken as the initial allocation for reaching the optimum allocation.

Let each of the rational number r (where r = =) of the ector $V = (IT,$ $E\, r$ = n,corresponding to agiven vector V, be nded up to the nearest integer 17 (17, is **either equal** to Tor T— 1, where° q, - 1 < r, 5_ q,) for s= 1,2,..., K to form the proportional allocation vector v, =) **with** some adjustments if needed to satisfy the restriction L n, = 17. **When r for** some s is less than unity, n, is taken as unity, since every stratum must have at least one psu in the sample

(4) The computational steps for the finding v_o

(a) Using the vector v_p (where $\sum n_s = n$) find $\max_s \frac{r_s}{n_s}$ and let for some $s = u$, $\frac{r_u}{n_u} = \max_s \frac{r_s}{n_s}$.

(b) If $\frac{r_s}{n_s - 1} \geq \frac{r_u}{n_u}$ for all s, then $v_o = v_p$

(c) If $\frac{r_s}{n_s - 1} < \frac{r_u}{n_u}$ for some $s = t$, say, but $t \neq u$; i.e., if $\frac{r_t}{n_t - 1} < \frac{r_u}{n_u}$, n_t is decreased to $n_t - 1$ and n_u is increased to $n_u + 1$. Then find revised $\max_s \frac{r_s}{n_s}$.

(d) If the condition in step (e) with the use of the revised maximum is satisfied for some other stratum, repeat step(c) for all strata. If this process is continued until a single stratum satisfying the condition is left we get a vo satisfying the condition in step (b) where v_p is changed to some other v .

(5) Illustration

1- In Table (1), corresponding to each s in col. (1), N_s is given in col.(2)and r_s is computed in col. (3) .The proportional allocation (n_s) is then found out in col. (4) of the table. According to step (a) of the procedure the ratio $\frac{r_s}{n_s}$ for each s is determined in col. (5) and the $\max_s \frac{r_s}{n_s}$ is obtained as 1.41 at $s=4$. It is shown that $\frac{r_s}{n_s - 1} = 1.11 < 1.41$. Hence $n_s = 6$ is decreased to $n_s - 1 = 5$ and $n_4 = 1$ is increased to $n_4 + 1 = 2$ and then $\frac{r_4}{n_4 + 1} = 0.70$. Now the max is 1.26 at s =7. Since $\frac{r_7}{n_7 - 1}$ =1.18 <1.26, n_7 is changed to

$n_7 - 1$ = 5, and n_{10} +1= 3 and $\frac{-}{3}$ = 0.84. Now max -1-1.18and

according to step (b) it is found that $\frac{r}{n_s - 1}$ '1.18 for all s. Hence,

v_0, whose element n_s is recorded for each s in Col. (6) of the table, is determined.

The minimum multiplier in case of v is 1.41 x $\frac{\sim}{-}$ = 200 and in case

of v, it is 1.18 x - =168. Hence, the multiplier corresponding to v_p

is 1.19 times that for v„.this means that if 100 ssu's are selected under v 119 ssu's will bt selected under v .

(1) TABLE

s	N	r	n	$\frac{r}{n_s}$		$\frac{r}{n_{0s}}$
1	228	1.60	2	0.80	2	0.80
2	246	1.73	2	0.86	2	0.86
3	790	5.56	6	0.93	5	1.11
4	200	1.41	1	1.41	2	0.70
5	320	2.25	2	1.12	2	1.12
6	926	6.53	6	1.09	6	1.09
7	358	2.52	2	1.26	3	0.84
8	543	3.82	4	0.96	4	0.96
9	276	1.94	2	0.97	2	0.97
10	839	5.90	6	0.98	5	1.18
11	389	2.74	3	0.91	3	0.91
Total	5115	36.00	36	-	36	-

2- In Table (2) max $\frac{-}{n}$ max $\frac{r}{n}$ = $\frac{1.41}{1.15}$ =1.23. So, if 100 ssu's are selected under v 123 ssu's will be selected under v.

TABLE (2)

s	N.	r	17	$\frac{r}{n}$		$\frac{r}{n_{0,}}$
1	413	2.60	3	0.87	3	0.87
2	1096	6.92	7	0.99	6	1.15
3	451	2.84	3	0.95	3	0.95
4	371	2.34	2	1.17	3	0.78
5	858	5.41	5	1.08	5	1.08
6	729	4.60	5	0.92	4	1.15
7	164	1.03	1	1.03	1	1.03
8	224	1.41	1	1.41	2	0.70
9	161	1.02	1	1.02	1	1.02
10	547	3.45	3	1.15	3	1.15
11	394	2.48	3	0.83	3	0.83
12	301	1.90	2	0.95	2	0.95
Total	5709	36.00	36	-	36	-

3- As $\max_{n,} \div \max_{n_{0,}} r = \frac{1.61}{1.28} = 1.26$ in table (3), if 100 ssu's are selected under v ,126 ssu's will be selected under v_0.

TABLE (3)

s	N	r	17	$\frac{r}{n}$		$\frac{r}{n_{0,}}$
1	36	0.94	1	0.94	1	0.94
2	62	1.61	1	1.61	2	0.80
3	103	2.68	3	0.89	3	0.89
4	42	1.09	1	1.09	1	1.09
5	43	1.12	1	1.12	1	1.12
6	147	3.83	4	0.96	3	1.28
7	28	0.73	1	0.73	1	0.73
Total	461	12.00	12	-	12	-

Table (4) has been set up for determining v_0. We find that

$\max_{n,}' + \max_{n_{0}}$ = is 1.03 in this case. This shows that with the increase in sample size of psu's the relative gain in number of ssu's due to v_0 is decreased.

TABLE (4)

s	N,	r	n	$\frac{r}{n}$	n_0	$\frac{r}{n_0}$
1	36	1.88	2	0.92	2	0.92
2	62	3.22	3	1.07	3	1.07
3	103	5.36	5	1.07	5	1.07
4	42	2.18	2	1.09	2	1.09
5	43	2.24	2	1.12	3	0.75
6	147	7.66	8	0.96	7	1.09
7	28	1.46	2	0.73	2	0.73
Total	461	24.00	24	-	24	-

Acknowledgment

The author is grateful to the referees for their comments on this paper.

References

1-Hansen MM., Hurwitz W.N and Madow, W.G. (1953): Sample Survey Methods and Theory, I, john Wiley & Sons, New York.

2- Lahore,D.B.(1954):Technical paper on some aspects of the development of the sample design. Sankhya, 14,264-316.

3- Cochran, W.G (1963): Sampling Techniques, (Second Edition) john Wiley & Sons, New York.

4- Murthy, M. N. (1967): Sampling Theory and Methods, Statistical Publishing Society, Calcutta.

5- Yamane, T. (1967). Elementary Sampling theory, Prentice - Hall, INC.Englewood cliffs, N.J.

6- Wynn, H.P. (1977). Minimax Purposive Survey Sampling Design. Journal of the American Statistical Association, vol..72.No.359.
7- Royall, R.M. & Cumberland, Vs'. G. (1978) Variance Estimation in Finite population Sampling. Journal of the American Statistical Association, vol .73.No.362.
8- Scheaffer,R,L. Mendenhall,W.& Ott,L. (1979). Elementary Survey Sampling. Second Edition. Wadsw orth Publishing Company, INC.
9-Scott, A.J. & Halt, D. (1982). The Effect of two - Stage Sampling on ordinary least squares Methods. Journal of the American Statistical Association, vol .77.No.380.
10- Jewell, N.P. (1985). Least Squares Regression with Data Arising from Stratified Sampling of the dents ariable. Biametrika. 72, 1, pp. 11-21.
11-Malec, D. & Sedransk, J. (1985). Bayesian Inference for Finite Population Parameters in Multistage cluster Sampling. Journal of the American Statistical Association, vol .80.No.392.
12- Bankier, M.D. (1986) Estimators Based on Several Stratified Samples with Application to Multiple Frame Surveys, Journal of the American Statistical Association, vol .81.No.396.
13-Valliant, R. (1987). Conditional Properties of Some Estimators in Stratified Sampling. Journal of the American Statistical Association, vol . 82.No.398.
14- Ghash, M, & Lahiri, P. (1987). Rabust Empirical Bayes Estimation of Means Frame Stratified Samples. . Journal of the American Statistical Association, vol .82. No. 400.
15-Christense
n, R. (1987). The Analysis of Two Stage Sampling Data by Ordinary Least Squares. Journal of the American Statistical Association, vol .81.No. 398.
16- Chaudhuri, A. & Stenger, H. 0992) Survey Sampling theory and Methods. Marcel Dekker, INC.