Jour. Ind. Soc. Ag. Statistics
Vol. XL, No. 2 (1988), pp. 83-95

SAMPLING FROM TWO-DIMENSIONAL POPULATIONS SPREAD OVER SPACE AND TIME

D. L. AHUJA AND A. K. SRIVASTAVA
Indian Agricultural Statistics Research Institute,
New Delhi-110012

(Received : April, 1984)

SUMMARY

The paper considers some aspects of sampling from two-dimensional populations spread over space and time. Combinations of equal probability sampling and systematic sampling over time dimension are considered. Further, a scheme with double sampling and systematic sampling has been examined.

Keywords : Aligned sampling, Two-dimensional population, Simple random sampling, Systematic sampling, Double sampling.

Introduction

Sampling in two-dimensional populations provides flexibility in the choice of sampling designs, as different sampling procedures may be used over the two-dimensions. In populations spread over space and time when a sampling unit in space needs to be observed over time, some sampling procedure has to be adopted for the time dimension taking into account the convenience of field work. Systematic sampling is a convenient procedure in such cases.

Sampling from two-dimensional populations has been studied in somewhat different contexts by Cochran [6], Quenouille [10], Das [7], Vos [11], Bellhouse and Rao [4], Bellhouse and Finlayson [5].

In the present paper, some operationally convenient procedures using different sampling methods over the two dimensions are investigated.

The problem is approached through varying probability sampling without assuming any super population model.

2. Sampling in Two Dimensions

Consider a population of NM units spread in two dimensions with N rows and M columns. For sampling nm units from the entire population, one alternative is to consider NM units as units of a single population from which a sample may be selected. Alternatively, samples may be selected from both the directions separately. The sampling procedures considered here are simple random sampling and systematic sampling in cases of aligned and unaligned selections. A selection is said to be aligned along rows (columns) if a number of rows (columns) are selected and selection in the other direction is confined to the selected rows (columns) only. The ultimate selection procedure leads to a varying probability sampling design and the usual estimation procedure for varying probabilities may be used.

2.1. *Notations*

In the subsequent notations, population values and sample values are respectively denoted by capital and small letters. Let y be the study character and Y_{ij} be the y-value for (i, j)th unit; $(i = 1, \ldots, N;$ $j = 1, \ldots, M)$; $Y_{i\cdot}$, $Y_{\cdot j}$ and Y are totals for rows, columns and the entire population respectively and $\bar{Y}_{i\cdot}$, $\bar{Y}_{\cdot j}$ and $\bar{Y}$ are the the corresponding means respectively. Let the population mean square be S^2 and mean squares for between and within row means be S^2_{br} and $\bar{S}^2_{wr}$ respectively. Corresponding mean squares for columns are S^2_{bc} and $\bar{S}^2_{wc}$.

For systematic samples with sampling interval k over time dimensions (columns) (assuming $M = mk$), define

$y_{iu\lambda}$: y-value for λth unit in the uth sample for ith row.

$\bar{y}_{iu}$: Sample mean for the uth sample corresponding to ith row.

$\bar{y}_{u}$: Sample mean of column means for the uth sample.

ρ_i : Intraclass correlation for the systematic samples in the ith row.

S^2_{syi} : $\dfrac{1}{k-1} \sum\limits_{u}^{k} (\bar{y}_{iu} - \bar{Y}_i)^2$

S^2_{sy} : $\frac{1}{k-1} \sum_{u}^{k} (\bar{y}_u - \bar{Y})^2$

S^2_{iu} : $\frac{1}{m-1} \sum_{\lambda}^{m} (y_{iu\lambda} - \bar{y}_{iu})^2$

: $\frac{1}{N} \sum_{i}^{N} S^2_{syi}$

π_{ij} : Probability of inclusion of (i, j)th unit in the sample.

$\pi_{ij, i'j'}$: Probability of inclusion (i, j)th unit and (i', j')th unit in the sample simultaneously.

$s_i s'_j$: A sampling procedure with a sampling methods s along rows and s' along the columns where i and j both can take the values 1 or 0 denoting respectively the alignment or unalignment along the respective dimensions.

The cases $s_1 s'_0$ and $s_0 s'$ are straightaway two stage sampling considering rows (columns) as primary stage units (psu's) and units within rows (columns) as secondary stage units (ssu's).

The above definition of $s_i s'_j$ for two dimensional populations is rather general and different sampling procedures may be adopted along the two dimensions. However in the present study we restrict to simple cases of equal probability sampling and systematic sampling for sampling along the two dimensions. Systematic sampling is considered in view of its practicability along time dimensions. Simple random sampling along space dimensions is used to illustrate the approach. Further extensions to more complex designs are rather straightforward in approach but more involved algebriacally.

2.2 *Equal Probability Sampling*

Denoting simple random sampling by "r" following procedures for sampling from two-dimensional populations follow:

(i) $r_0 r_0$, (ii) $r_1 r_0$, (iii) $r_0 r_1$ and (iv) $r_1 r_1$

Evidently, $r_0 r_0$ is selection of nm units from NM units by simple random sampling, $r_1 r_0$ and $r_0 r_1$ are two-stage sampling with simple random sampling at both the stages.

In the case of $r_1\,r_1$ with two-way alignment, samples of n rows and m columns are selected from N rows and M columns respectively by simple random sampling without replacement. Thus, we get nm ultimate units corresponding to nm points of intersection of the selected rows and columns.

The Horvitz-Thompson (HT) estimator of the population total is

$$\hat{Y}_{HT} = \frac{MN}{mn} \sum_{i=1}^{n} \sum_{j=1}^{m} y_{ij} = NM\,\bar{y}_{nm}. \qquad (1)$$

A pair of units in this case either belong to (i) same row, (ii) same column or (iii) neither same row nor same columns.

Correspondingly, the pairwise inclusion probabilities are given by

$$\pi_{ij,ij'} = \frac{nm\,(m-1)}{NM\,(M-1)} \quad (j \neq j'\,;\, i = 1, \ldots, N)$$

$$\pi_{ij,\,i'j} = \frac{nm\,(n-1)}{NM\,(N-1)} \quad (i \neq i',\, j = 1, \ \ldots, M)$$

$$\pi_{ij,\,i'j'} = \frac{nm\,(n-1)\,(m-1)}{NM\,(N-1)\,(M-1)} \quad \begin{bmatrix} i \neq i' = 1, \ldots, N \\ j \neq j' = 1, \ldots, M \end{bmatrix}$$

The variance of $\hat{Y}_{HT}$ is given by

$$V(\hat{Y}_{HT}) = \frac{1}{2} \sum_{ij \neq i'j'} \sum (\pi_{ij}\ \pi_{i'j'} - \pi_{ij,\,i'j'}) \left\{ \frac{y_{ij}}{\pi_{ij}} - \frac{y_{i'j'}}{\pi_{i'j'}} \right\}$$

which after a little simplification, gives

$$V(\hat{Y}_{HT}) = V_{r_1 r_1} = N^2M^2 \Bigg[\left(\frac{1}{m} - \frac{1}{M} \right) S^2_{oc} + \frac{1}{m} \left(\frac{1}{n} - \frac{1}{N} \right) S_{wc} + \frac{M\,(m-1)\,(N-n)}{Nmn\,(M-1)} \left(S^2_{br} - \frac{\bar{S}^2_{wc}}{M} \right) \Bigg]. \qquad (2)$$

The corresponding variance for $r_1 r_0$ may be simplified to

$$V_{r_1 r_0} = N^2 M^2 \left[\left(\frac{1}{n} - \frac{1}{N} \right) S^2_{br} + \frac{1}{n} \left(\frac{1}{m} - \frac{1}{M} \right) \bar{S}^2_{wr} \right].$$

The effect of column alignment in the presence of row alignment may be given by

$$V_{r_1 r_1} - V_{r_1 r_0} = \frac{N^3 M\,(n-1)\,(M-m)}{nm\,(N-1)} \left(S^2_{bc} - \frac{\overline{S}^2_{wr}}{N} \right). \qquad (3)$$

When units within the columns are obtained randomly, the effect of alignment reduces to zero.

2.3 *Combination of Equal Probability Sampling and Systematic Sampling*

For sampling from two-dimensional populations spread over space and time, cnnsider equal probability sampling over space and systematic sampling over time. Further, it is operationally convenient to observe the selected space units on the same selected days. This leads to aligned systematic sampling over time. Denoting systematic sampling method of selection by "s_y" the procedure $r_0 s_{y_1}$ is clearly two-stage sampling with time units as psu's.

In the procedure $r_1\, s_{y1}$, n rows corresponding to n space sample units are selected by SRSWOR while m columns standing for m days are selected by systematic sampling. A total of nm units corresponding to the intersection of the selected rows and columns form the sample. The inclusion probability for (i, j)th unit is

$$\pi_{ij} = \frac{nm}{NM}\ ; \quad i = 1, \ldots, N, \qquad j = 1, \ldots, M.$$

The pairwise inclusion probabilities are given as follows :

$$\pi_{ij,\, ij'} = \frac{nm}{NM}\ ; \quad \text{when } j \text{ and } j' \text{ belong to same systematic sample; } i = 1, \ldots, N$$

$(j \neq j')$

$$= 0 \qquad \text{otherwise}$$

$$\pi_{ij,\, i'j} = \frac{nm\,(n-1)}{NM\,(N-1)}\ ; \qquad j = 1, \ldots, M$$

$(i \neq i')$

$$\pi_{ij,\, i'j'} = \frac{nm\,(n-1)}{NM\,(N-1)} \quad \text{when } j \text{ and } j' \text{ belong to same systematic sample, } i \neq i' = 1, \ldots, N$$

$$= 0 \qquad \text{otherwise}$$

The Horvitz-Thompson estimator of the population total remains same as in (1) and its variance is given by

$$V_{r1\,sv_1} = V\,(\hat{Y}_{HT}) = \frac{N^2M^2}{2n^2m^2}\Big[\sum_{i}^{N}\sum_{j\neq j'}^{M}\sum (\pi_{ij}\cdot\pi_{ij'} - \pi_{ij,\,ij'})\,(y_{ij} - y_{ij'})^2$$

$$+ \sum_{j}^{M}\sum_{i\neq i'}^{N}\sum (\pi_{ij}\,\pi_{i'j} - \pi_{ij,\,i'j})\,(y_{ij} - y_{i'j})^2$$

$$+ \sum_{i\neq i'}^{N}\sum\sum_{j\neq j'}^{M}\sum (\pi_{ij}\,\pi_{i'j'} - \pi_{ij,\,i'j'})\,(y_{ij} - y_{i'j'})^2\Big]$$

which further simplifies to

$$V_{r1sv1} = \frac{N^2M^2}{2n^2m^2}\Big[\frac{2m^2n\,(MN-1)\,(N-n)}{MN\,(N-1)}\,S^2$$

$$+ \frac{2n\,(n-1)\,m^2N}{M\,(N-1)}\sum_{u}^{k}(\bar{y}_u - \bar{Y})^2$$

$$+ \frac{2m^2n\,(M-N)\,(n-N)}{M\,(N-1)}\,\bar{S}^2_{wr}$$

$$+ \frac{2m^2n\,(k-1)\,(N-n)}{MN\,(N-1)}\sum_{i}^{N} S^2_{syi}\Big].$$

Using the identity

$$(MN-1)\,S^2 = M\,(N-1)\,S^2_{br} + N\,(M-1)\,\bar{S}^2_{wr},$$

V_{r1sv1} reduces to

$$V_{r1sv1} = N^2M^2\Big[\left(\frac{1}{n} - \frac{1}{N}\right)S^2_{br}$$

$$+ \frac{N\,(k-1)}{k\,(N-1)}\left\{\left(\frac{1}{n} - \frac{1}{N}\right)\bar{S}^2_{sy} + \frac{n-1}{n}\,S^2_{sy}\right\}\Big].$$

Further, defining $\bar{\rho}_w \bar{S}_{wr}$ as $\sum_i^N \rho_i S_{ir}^2 / N$, it follows that

$$\bar{S}_{sy}^2 = \frac{1}{N} \sum_i^N S_{syi}^2 = \frac{(km-1)}{m^2(k-1)} \left[\bar{S}_{wr}^2 + (m-1) \frac{1}{N} \sum_i^N \rho_i S_{ir}^2 \right]$$

$$= \frac{(km-1)}{m^2(k-1)} \bar{S}_{wr}^2 [1 + (m-1) \bar{\rho}_w].$$

Also,

$$S_{sy}^2 = \frac{km-1}{m^2(k-1)} S_{bc}^2 [1 + (m-1) \rho_b]$$

where ρ_b is intraclass correlation of column means for systematic samples. Substituting for $\bar{S}_{sy}^2$ and S_{sy}^2 we get

$$V_{r_1 sy_1} = \left(\frac{1}{n} - \frac{1}{N} \right) \left[S_{br}^2 + \frac{N(M-1)}{mM(N-1)} \bar{S}_{wr}^2 [1 + (m-1) \bar{\rho}_w] \right.$$
$$\left. + \frac{N(M-1)}{nM(N-1)} \frac{(n-1)}{n} S_{bc}^2 [1 + (m-1) \rho_b.] \right] \quad (5)$$

When $k = 1$, the sampling reduces to one-dimensional random sampling. The variance expression also reduces to corresponding situation. Similarly, when $n = N$, there is only systematic sampling over time and N units belonging to columns may be considered as clusters. The corresponding variance formula follows from the above expression. The variance of HT-estimator corresponding to r_{0sy1} is

$$V_{r_0 sy_1} = V(\hat{Y}) = N^2 M^2 \left[\frac{k-1}{k} S_{sy}^2 + \frac{1}{m} \left(\frac{1}{n} - \frac{1}{N} \right) \bar{S}_{wc}^2 \right]. \quad (6)$$

Thus, alignment effect along space dimension, in the presence of time alignment, is

$$V_{r_1 sy_1} - V_{r_0 sy_1} = \left(\frac{1}{n} - \frac{1}{N} \right) \left(S_{br}^2 - \frac{S_{wc}^2}{M} \right)$$
$$+ \left(\frac{1}{n} - \frac{1}{N} \right) \left(\frac{k-1}{k} \right) \left[\frac{N}{N-1} \left(\bar{S}_{sy}^2 - S_{sy}^2 \right) - \frac{\bar{S}_{wc}^2}{m} \right]. \quad (7)$$

This alignment effect has two components. The first component is due to random aligned sampling and additional second component is due to introduction of systematic sampling.

3. $r_1 s y_1$ Procedure with Double Sampling

Use of double sampling is sometimes considered suitable for studying two-dimensional populations particularly when the study character (say, production) aggregated over entire time period is highly correlated with the corresponding production for some of the time units (Kaistha and Goel [9]). The production of the specific time period serves as auxiliary character and is observed on a larger sample of space units. The total production over the entire time span is the corresponding study character. A combination of $r_1 s y_1$ with double sampling is studied here.

3.1. *The Sampling Procedure*

Consider a two-dimensional population with N space units (units) each spread over M time units (intervals). The sampling procedure consists of selecting n' first phase units which are observed for an interval chosen on the basis of correlation coefficients of the interval-wise yields with the total plot yields. Select a sub-sample of n units from the n' units by SRSWOR which are observed systematically over time, using $r_1 s y_1$ procedure.

The suffix (j) denotes the jth time period whose character value is to be used as auxiliary character on the basis of its high correlation coefficient (ρ) with the corresponding value for the entire time period. The population mean $\bar{Y}_{(j)}$ and sample means $\bar{y}_{n'(j)}$ and $\bar{y}_{n\,(j)}$ are accordingly defined. The mean squares and mean sum of products for the population and samples are defined as follows

$$S_{br\,(j)} = \frac{1}{N-1} \sum_{i}^{N} (Y_{ij} - \bar{Y}_{N(j)})\,(\bar{Y}_{i.} - \bar{Y})$$

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

$$s_{br(j)} = \frac{1}{n-1} \sum_{i}^{n} (y_{ij} - \bar{y}_{n(j)})\,(\bar{y}_{i.} - \bar{y}_{nm})$$

$$s^2_{(j)} = \frac{1}{n-1} \sum_i^n (y_{ij} - \bar{y}_{n(j)})^2$$

and

$$\bar{y}_{nm} = \text{Mean of the systematic sample of } nm \text{ units}$$

$$= \frac{1}{n} \sum_i^n \frac{1}{m} \sum_j^m y_{ij} = \frac{1}{n} \sum_i^n \bar{y}_{i.} \, .$$

3.2. *Estimation*

An estimator of population mean $\bar{Y}$ is given by

$$\hat{\bar{Y}}_{dsy} = \bar{y}_{nm} + \hat{\beta}\,(\bar{y}_{n'(j)} - \bar{y}_{n(j)})$$

where

$$\hat{\beta} = \frac{s_{br(j)}}{s^2_{(j)}} \text{ is an estimator of } \beta = \frac{S_{br(j)}}{S^2_{(j)}} \, .$$

As in the case of usual double sampling technique with the assumption that the effect of estimating β is small enough, the variance of $\hat{\bar{Y}}_{dsy}$ may be obtained as

$$V_{dsy} = V_{r1sy1} - \left(\frac{1}{n} - \frac{1}{n'}\right) \rho^2 S^2_{br}$$

$$= \frac{\rho^2 S^2_{br}}{n'} + \frac{1}{n}\Big[[\,(1-\rho^2)\, S^2_{br} + \frac{k-1}{k}\,(\bar{S}^2_{sy} - S^2_{sy})$$

$$+ \frac{k-1}{k}\, S^2_{sy}\Big]$$

$$= V' + \left(\frac{k-1}{k}\right) S^2_{sy}$$

where

$$V' = \frac{\rho_2 S^2_{br}}{n'} + \frac{1}{n}\Big[(1-\rho^2)\, S^2_{br} + \frac{k-1}{k}\, S^2_{sy} - S^2_{sy}\Big]$$

$$= \frac{A}{n'} + \frac{B}{n}$$

where

$$A = \rho^2 S^2_{br}$$

$$B = (1-\rho^2)\, S^2_{br} + \frac{k-1}{k}\,(\bar{S}^2_{sy} - S^2_{sy}).$$

3.3 *Efficiency of the Suggested Procedure*

Let C_0 be the total cost and c be the cost for observing one space-time unit. The cost function in the present case may be written as

$$C_0 = C_n' + cnm = C_1 n' + C_2 n$$

where

$$C_1 = c,\ C_2 = cm = \frac{cM}{k}.$$

Minimisation of V' with respect to n' and n for given C_0 leads to

$$V'_{\min} = \frac{\sqrt{AC_1} + \sqrt{BC_2})}{C_0}.$$

For the fixed cost C_0, the variance in the usual SRSWOR procedure is given by

$$V_2 = \frac{cM}{C_0} S_{br}^2.$$

Thus the relative efficiency of the proposed procedure is given by

$$\frac{V_2}{V_{dsy}} = \frac{(c/C_0)\ MS_{br}^2}{V'_{\min} + \frac{k-1}{k} S_{sy}^2}$$

$$= \left[\left[\frac{\rho}{\sqrt{M}} + \sqrt{\frac{1}{k}\left[(1-\rho^2) + \frac{k-1}{kM}\frac{(\bar{S}_{sy}^2 - S_{sy}^2)}{S_{br}^2}\right]}\right]^2 + \frac{(k-1)}{kM}\frac{C_0 S_{sy}}{cS_{br}^2}\right]^{-1}$$

The percentage relative gain in efficiency (G) due to proposed procedure is given by

$$G = 100\left[\frac{V_2}{V_{dsy}} - 1\right].$$

TABLE 1 — PERCENTAGE GAIN IN EFFICIENCY (G) DUE TO PROPOSED PROCEDURE (DOUBLE SAMPLING — $r_1 s y_1$) OVER USUAL PROCEDURE

ρ	*5-day interval* ($M = 18$) G, K = 2	G, K = 3	G, K = 4	G (D)	*7-day interval* ($M = 13$) G, K = 2	G, K = 3	G, K = 4	G (D)	*10-day interval* ($M = 10$) G, K = 2	G, K = 3	G, K = 4	G (D)
0.50	86.63	158.16	225.05	8.63	77.50	142.48	200.96	6.12	67.26	129.07	181.44	4.19
0.55	91.81	163.09	229.19	12.87	81.33	145.38	202.41	9.66	69.84	130.40	180.88	7.11
0.60	98.90	170.33	236.03	18.44	86.23	150.28	206.13	14.38	73.87	133.48	182.31	11.11
0.65	108.46	180.47	246.13	25.76	94.46	157.61	212.55	20.65	79.68	138.67	186.05	16.50
0.70	121.35	194.43	260.52	35.52	104.92	168.11	222.37	29.05	87.83	146.55	192.62	23.77
0.75	139.01	213.74	280.78	48.87	119.40	182.94	236.77	40.55	99.24	158.05	202.90	33.75
0.80	164.03	241.07	309.72	67.94	140.00	204.18	257.76	56.90	115.51	174.82	218.42	47.93
0.85	201.63	281.66	352.77	97.10	179.96	235.85	289.22	81.70	139.79	200.04	242.06	69.30
0.90	264.32	347.59	422.30	147.42	222.38	287.13	339.78	123.83	179.42	240.94	280.16	105.81

An Illustration

In order to investigate the relative performance of 'Double Sampling—r_1sy_1' in comparison to the double sampling alone, the data regarding yield of vegetable from a vegetable survey, for estimating cost of cultivation of important vegetable crops in Delhi (1978-81) has been used. We confine to data for tomato crop for the year 1979-80. The results are presented in Table 1. The gains in efficiency, G are presented for ρ ranging between 0.50 to 0.90 and $k = 2, 3$ and 4. The variation of G for different values of C_0/c was found to be negligible. Therefore, results are presented for a particular value as $(C_0/cNM) = 0.05$. The gain in efficiency due to double sampling alone is also presented in the table as $G(D)$. Intervals of 5, 7 and 10 days are considered as time units.

The performance of proposed procedure appears to be convincingly evident. It is also seen that the gains due to double sampling only ($G(D)$) are quite small as compared to the proposed procedure particularly for smaller values of ρ.

G increases with ρ, for smaller values of ρ. The contribution of systematic sampling appears to be dominant over that of double sampling. The gains due to application of double sampling in combination with systematic sampling appear to be large only for larger values of ρ ($\rho \geqslant 0.8$).

REFERENCES

[1] Ahuja, D.L. (1983): Some studies in sampling from two dimensional populations, *Ph.D. thesis*, IARI. New Delhi.

[2] Bellhouse, D.R. (1977): Some optimal designs for sampling in two dimensions, *Biometrika* 64 : 605-611.

[3] Bellhouse, D.R. (1981): Spatial sampling in the presence of trend. *JSPI* 5 (4): 365-375.

[4] Bellhouse, D. R. and J. N. K. Rao (1975) : Systematic sampling in the presence of a trend, *Biometrika* 62 : 694-697.

[5] Bellhouse, D. R. and W. D. Finlayson (1979) : An empirical study of probability sampling designs, *Canad. J. Archaeology* 105-123.

[6] Cochran, W. G. (1946) : Relative accuracy of systematic stratified random samples for a certain class of populations, *Ann. Math. Statist.* 17 : 164-177.

[7] Das, A.C. (1950) : Two-dimensional systematic sampling and associated stratified and random sampling, *Sankhya* 10 : 95-108.

[8] Horvitz, D. G. and Thompson, D. J. (1952) : A generalisation of sampling without replacement from a finite universe, *Jour. Amer. Stat. Assoc.* 47 : 663-685.

[9] Kaistha, A. C. and Goel, B. B. P. S. (1981) : Estimation of vegetable production using partial harvest data, *Jour. Ind. Soc. Agr. Stat.* **33**(1) : 110-118.

[10] Quenouille, M. H. (1949) : Problems in a plane sampling, *Ann. Math. Stat.* **20** : 355-375.

[11] Vos, J. W. E. (1964) : Sampling in space and time, *Rev. Int. Statist. Inst.* **32** : 226-275.