The Annals of Mathematical Statistics
1972, Vol. 43, No. 6, 2031-2034

SUCCESSIVE SAMPLING WITH p ($p \geqq 1$) AUXILIARY VARIABLES[1]

BY A. R. SEN[2]

Department of Indian Affairs and Northern Development Canada

In succesive sampling on two occasions, the theory developed so far aims at providing the optimum estimate by combining (i) a double-sampling regression estimate from the matched portion of the sample and (ii) a simple mean based on a random sample from the unmatched portion of the sample on the second occasion. Theory has been generalized in the present note by using a double-sampling multivariate ratio estimate using p auxiliary variates ($p \geqq 1$) from the matched portion of the sample. Expressions for optimum matching fraction and of the combined estimate and its error have been derived and results are presented for some special cases which have practical applications.

1. Summary. Theory has been generalized in the present note to provide the optimum estimate in sampling on two occasions by using a double-sampling multivariate ratio estimate using p auxiliary variables ($p \geqq 1$) from the matched portion of the sample.

2. Introduction. In successive sampling on two occations, the theory developed so far (Jessen (1942); Patterson (1950)) aims at providing the optimum estimate by combining (i) a double-sampling regression estimate from the matched portion of the sample where the *large sample* is the first sample and the auxiliary variable X is the value of Y on the first occasion and (ii) a simple mean based on a random sample from the unmatched portion on the second occation. The present note utilizes the information from the first occasion on p ($p \geqq 1$) auxiliary variables $X_1, \cdots, X_p$ (whose population means "$\bar{X}_1$", "$\bar{X}_2$", $\cdots$, "$\bar{X}_p$" are unknown) to provide the optimum estimate of the parameter mean "$\bar{Y}$" where the variable under study Y refers to the second occation; one of the auxiliary variables may be the value of Y on the previous occasion.

3. Theory. Suppose that simple random samples of size n are drawn on both occasions from a population which is composed of the same set of N elements on the two occasions. Also, suppose $n - u$ of the units in the sample selected on first occasion are retained and the remaining u units (u for unmatched) are replaced by a new selection. Let $X_1, \cdots, X_p$ be the values of the p auxiliary variables on the first occasion on a matched unit for which the value on the second occasion is Y, also, let the population variance S^2 of Y be the same on both occasions.

Received June 10, 1971; revised April 4, 1972.

[1] Presented at the I.M.S. European Meeting held in Hannover, Germany, August 19-26, 1970.

[2] Now with the Department of the Environment, Canada.

Following Olkin (1958) and Goswami and Sukhatme (1965) the multivariate double-sampling ratio estimate $\bar{Y}_R$ from the matched portion is

$$(1)\qquad \bar{Y}_R = \sum_{i=1}^{p} W_i \frac{\bar{Y}}{\bar{X}_i} \bar{X}_i' \equiv \sum_{i=1}^{p} W_i R_i \bar{X}_i' .$$

where $\bar{Y}$ is the sample mean of the matched portion on the second occasion, $\bar{X}_i$ is the estimate for the same sample on the ith auxiliary variable based on $n - u$ units, $\bar{X}_i'$ the corresponding estimate based on the large sample of size n where the weights W_i ($\sum W_i = 1$) are to be determined to maximize the precision of $\bar{Y}_R$. It is well known that the estimator (1) will be subject to bias which may be considered negligible for moderately large sample size.

Now,

$$(2)\qquad V(\bar{Y}_R) = \sum_i^p W_i^2 V_{ii} + 2 \sum \sum_{i<j} W_i W_j V_{ij} ,$$

where

$$V_{ii} = V\left(\frac{\bar{Y}}{\bar{X}_i} \cdot \bar{X}_i'\right) \quad \text{and} \quad V_{ij} = \mathrm{Cov}\left(\frac{\bar{Y}}{\bar{X}_i} \bar{X}_i', \frac{\bar{Y}}{\bar{X}_j} \bar{X}_j'\right).$$

Equation (2) can be written as

$$(3)\qquad V(\bar{Y}_R) \doteq (A/n - u) + (B/n) ,$$

where

$$A = \sum_{i,j}^{p} W_i W_j S_{b_{ij}}^2 ; \qquad S_{b_{ij}}^2 = S^2 - R_i S_{yx_i} - R_j S_{yx_j} + R_i R_j S_{x_i x_j} ;$$
$$B = S^2 - A .$$

S_{yx_i}, S_{yx_j}, $S_{x_i x_j}$ have their usual meanings (Cochran, (1963)). For optimum weight function W_i and $V_{\mathrm{opt}}(\bar{Y}_R)$; see Olkin (1958).

If $\bar{Y}_u$ denotes the mean per unit estimate from the unmatched portion u, and $\bar{Y}'$ the weighted mean of $\bar{Y}_{R\,\mathrm{opt}}$ and $\bar{Y}_u$ where $\bar{Y}_{R\,\mathrm{opt}}$ is the optimum value of $\bar{Y}_R$ from (1), we have

$$(4)\qquad \bar{Y}' = \lambda_1 Y_{R\,\mathrm{opt}} + (1 - \lambda_1)\bar{Y}_u ,$$

where

$$(5)\qquad V(\bar{Y}') = \lambda_1^2 V(\bar{Y}_{R\,\mathrm{opt}}) + (1 - \lambda_1)^2 V(\bar{Y}_u) ,$$

since

$$\mathrm{Cov}(\bar{Y}_{R\,\mathrm{opt}}, \bar{Y}_u) = 0 .$$

The value of λ_1 which minimizes $V(\bar{Y}')$ will be given by

$$\lambda_1 = \frac{V(\bar{Y}_u)}{V(\bar{Y}_{R\,\mathrm{opt}}) + V(\bar{Y}_u)} .$$

Hence, substituting the value of λ_1 in (5) we have

$$(6)\qquad V_{\mathrm{opt}}(\bar{Y}') \doteq \frac{(A'/(n - u) + B'/n)(S^2/u)}{A'/(n - u) + B'/n + S^2/u} ,$$

where A', B' denote the values of A and B in (3) obtained by substituting the optimum values of W_i and W_j.

The optimum value of u will be obtained by minimizing (6) with respect to variation in u and is given by the solution of the equation

$$B'\left(\frac{u}{n}\right)^2 - 2(A' + B')\left(\frac{u}{n}\right) + \left\{\frac{(A' + B')^2 - A'S^2}{B'}\right\} = 0 . \tag{7}$$

The value of u given by the solution of

$$V(\bar{Y}_{R\,\mathrm{opt}}) = V(\bar{Y}_u) \tag{8}$$

reduces to

$$B'(u/n)^2 - (S^2 + A' + B')(u/n) + S^2 = 0 . \tag{9}$$

Equation (9) is identically the same as (7) since $A' + B' = S^2$ by (3). Hence, the optimum value of u which minimizes (6) is given by the solution of (8). Hence, (6) reduces to

$$S^2/2\hat{u} , \tag{10}$$

where $\hat{u}$ is given by the solution of (7) and $n - \hat{u}$ denotes the optimum matching.

4. Special case. For the special case

$$\rho_{ij} = \rho_1 \qquad (i \neq j);\ (i, j = 1, \cdots, p) \text{ and}$$
$$\rho_{0i} = \rho_0 ,$$

where ρ_{ij} is the correlation between X_i and X_j and ρ_{0i} is the correlation between Y and X_i, it has been shown by Khan and Tripathi (1967) that $V(\bar{Y}_R)$ using p-auxiliary variables is given by

$$\frac{S^2}{p}\left\{\frac{p^+(p-1)\rho_0\rho_1^2 - (2p-1)\rho_1^2}{n-u} + \frac{(2p-1)\rho_1^2 - (p-1)\rho_0\rho_1^2}{n}\right\} . \tag{11}$$

Hence

$$V_{\mathrm{opt}}(\bar{Y}') = S^2/2u ,$$

where u is given by

$$u/n = \frac{p \pm [p^2 - p\{\rho_1^2(2p-1) - \rho_0\rho_1^2(p-1)\}]^{\frac{1}{2}}}{\rho^2\{(2p-1) - \rho_0(p-1)\}} . \tag{12}$$

A particular case of special interest is the theory for two auxiliary variables (X_1, X_2; $p = 2$) which has the frequent application

$$V_{\mathrm{opt}}(\bar{Y}') = S^2/2u ,$$

where

$$u/n = \frac{2 \pm [4 - 2\rho_1^2(3 - \rho_0)]^{\frac{1}{2}}}{\rho_1^2(3 - \rho_0)} . \tag{13}$$

REFERENCES

[1] COCHRAN, W. G. (1963). *Sampling Techniques*, 2nd ed. Wiley, New York.

[2] GOSWAMI, J. N. and SUKHATME, B. V. (1965). Ratio method of estimation in multi-phase sampling with several auxiliary variables. *J. Indian Soc. Agric. Statist.* **17** 83–103.

[3] JESSEN, R. J. (1942). Statistical investigation of a sample survey for obtaining farm facts. *Iowa Agricultural Experimental Station Research Bulletin* 304.

[4] KHAN, S. and TRIPATHI, T. P. (1967). The use of multivariate auxiliary information in double-sampling. *J. Indian Statist. Assoc.* **5** 42–48.

[5] OLKIN, INGRAM (1958). Multivariate ratio estimation for finite populations. *Biometrika* **45** 154–65.

[6] PATTERSON, H. D. (1950). Sampling on successive occasions with partial replacement of units. *J. Roy. Statist. Soc. Ser. B* 241–255.

DEPARTMENT OF ENVIRONMENT
400 LAURIER AVENUE WEST
OTTAWA, ONTARIO K1A OH4, CANADA