A FURTHER CONTRIBUTION TO THE THEORY OF UNIVARIATE SAMPLING ON SUCCESSIVE OCCASIONS

By B. D. Tikkiwal

Institute of Statistics, Raleigh, N.C., U.S.A. and Karnatak University, Dharwar, India

1. Introduction

The general theory of Sampling on Successive Occasions for a single variate has been studied by Patterson[2] and the author.[3] Both the authors independently arrived at the same minimum-variance linear unbiased estimator of the population mean on the hth occasion under a specified sampling scheme and the pattern of correlation for the h occasions.

The pattern of correlation assumed by the author is slightly more general than that used by Patterson for the exact theory. The correlation between the units on the consecutive occasions is allowed to vary, but the correlation between the units two or more than two occasions apart is assumed to be the product of the correlations between the units on all pairs of consecutive occasions occurring between these two occasions.

The variance of the estimator on the hth occasion was also obtained by both the authors under the assumption that the various correlation and regression coefficients occurring in the estimator are known in advance but not calculated from the sample. Later Narain[1] suggested a modification in the weights assigned by the authors to the two estimators available on the hth occasion in order to get the final estimator on that occasion.

The purpose of this paper is to obtain the variance of the minimum-variance linear unbiased estimator of the population mean on the hth occasion, when the correlation and regression coefficients occurring in the estimator are estimated from the common units in the sample, and then give an estimator of the variance itself. Finally the effect of the modification in weights suggested by Narain is examined in Section 4.

2. Successive Univariate Sampling

In this section we shall summarise the earlier results as developed by the author for further reference in the paper. Suppose we are interested in studying the variate Y of an infinite population π over a period of time. The purpose of the study is to estimate the population mean of the variate on each of the h occasions over the specified period. For this, the sample of size n_r on the rth occasion consists of two parts: (1) The units n_r' which are also observed for the same variate at least on the previous occasion. (2) The new units n_r'' from the population units not selected so far in the sample. Thus $n_r = n_r' + n_r''$. The correlation between the units on the ith and jth occasions is assumed to be $\pi_{t=i}^{j-1} \rho_{t,\, t+1}$, where $\rho_{t.\, t+1}$ is the correlation between the units on tth and $(t+1)$th occasions.

The minimum-variance unbiased linear estimator Y_h, under the assumed pattern of correlation, of the population mean M_h on the hth occasion is given by

$$Y_h = (1 - \phi_h)\, Z_h + \phi_h \bar{y}_h'' \tag{1}$$

where

$Z_h \quad = \bar{y}_h' + \beta_{h-1,\, h}\, (Y_{h-1} - \bar{x}_h')$.

$\bar{y}_h' \quad$ = mean on the hth occasion of the n_h' units also observed on the $(h-1)$th occasion.

$\bar{x}_h' \quad$ = mean on the preceding $(h-1)$th occasion of the same n_h' units.

$\bar{y}_h'' \quad$ = mean on the hth occasion of the n_h'' units observed for the first time on this occasion.

$\beta_{h-1,\, h} = \rho_{h-1,h} \dfrac{\sigma_h}{\sigma_{h-1}}$, σ_t^2 for $t = 1, 2, \cdots h$ is the variance of the variate values on the tth occasion.

The weighting coefficient ϕ_h is given by

$$\frac{\phi_h}{1-\phi_h} = \frac{n_h''}{n_h'}(1 - \rho^2_{h-1,\, h}) + \rho^2_{h-1,\, h}\, \frac{n_h''}{n''_{h-1}}\, \phi_{h-1} \tag{2}$$

and

$$\phi_1 = \frac{n_1''}{n_1},$$

where n_1 is the total sample size on the first occasion and $n_1'' = n_1 - n_2'$.

If $\beta_{h-1,\, h}$ is estimated from the n_h' common units in the sample, Narain[1] suggests the use of modified weight ϕ_h' obtained by replacing $\rho^2_{h-1,\, h}$ by

$$\left(\rho^2_{h-1,\, h} - \frac{1 - \rho^2_{h-1,\, h}}{n_h' - 3}\right)$$

in the recurrence formula (2). The correlation coefficient $\rho_{t-1,\,t}$ for $t = 2, \cdots h$ will also be estimated in general from n_t' common units.

When ϕ_t and $\beta_{t-1,\,t}$ for $t = 2, \cdots h$ are known in advance but not calculated from the sample, then

$$\text{Var}\,(Y_h) = \phi_h \frac{\sigma_h^2}{n_h''} \tag{3}$$

as given in [2] and [3].

3. Modified Formula for the Variance and its Sample Estimator

Suppose the various regression and correlation coefficients occurring in Y_h are now estimated from the common units, then the variance of Y_h is given by

$$\text{Var}\,(Y_h) = \frac{\sigma_h^2}{n_h''} E\,(\hat{\phi}_h) \tag{4}$$

$\hat{\phi}_h$ being the sample estimate of ϕ_h. This is so, provided n_t' common units on the tth occasion forms a sub-sample of n''_{t-1} new units on $(t-1)$th occasion for $t > 2$. If this conditions is not satisfied and if the various correlation and regression coefficients are still estimated from the common units then Y_h will be a consistent and asymptotically unbiased estimator of M_h the population mean on the hth occasion and its mean square deviation V will now satisfy the inequality.

$$\phi_h \frac{\sigma_h^2}{n_h''} \leqslant V \leqslant \frac{\sigma_h^2}{n_h''} E\,(\hat{\phi}_h) \tag{5}$$

A consistent and asymptotically unbiased estimator $\hat{V}$ of the variance or the mean square deviation of Y_h as the case may be is given by

$$\hat{V} = \hat{\phi}_h \frac{\hat{\sigma}_h^2}{n_h''} \tag{6}$$

Where, $\hat{\sigma}_h^2$ is the usual estimator of σ_h^2 based on n_h units.

Proof.—When

$$h = 2,$$

$$E\,(Y_2 - M_2) = E_1\, E_2\, (Y_2 - M_2)$$

Where E_2 is the conditional expectation when the sample correlation and regression coefficients r_{12} and b_{12} are held constant. The distribution of the various sample means occurring in Y_2 are independent of the distributions of r_{12} and b_{12} under normality assumption. Therefore, the conditional expectation of Y_2 is the same as the unconditional one when r_{12} and b_{12} are some given constants. But the unconditional expectation of $(Y_2 - M_2)$ is zero and so the estimator is unbiased. Also,

$$\text{Var}\,(Y_2) = E_1 E_2 (Y_2 - M_2)^2$$

The unconditional expectation of $(Y_2 - M_2)^2$, when r_{12} and b_{12} are some given constants, is $\hat{\phi}_2 \sigma_2{}^2/n_2''$ [*cf.* Eq. (3)] and hence the equation (4) for $h = 2$. For $h > 2$, when n_t' is the sub-sample of $n_t''_{-1}$ for $t > 2$, the various sample means occurring in Y_h are still independent of $r_{t-1,\,t}$ and $b_{t-1,\,t}$ for $t = 2, \cdots h$ and so the estimator Y_h is still unbiased and the variance of the estimator is given by equation (4).

If n_t' do not satisfy the above restriction, the various regression and correlation coefficients may still be estimated from those parts of common units which are sub-samples of the new units on the previous occasions and all the results will hold good. But one can use better estimates of regression and correlation coefficients. One such set of estimates will be that utilising whole of the common units on all the h occasions. In this case, Y_h is a consistent estimator of the population mean on the hth occasion and its bias tends to zero with increasing sample sizes on the h occasions, *i.e.*, it is asymptotically unbiased. For, the estimator Y_h is a function of various sample means and regression and correlation coefficients say r in number. We expand this function by the following Taylor's formula

$$\begin{aligned} f(x_1, x_2, \cdots x_r) &= f(a_1, a_2, \cdots a_r) \\ &\quad + \Sigma^r_{i=1} \left[(x_i - a_i) \frac{\partial f(A_1, A_2, \cdots . A_r)}{\partial A_i} \right], \\ &\quad A_i = a_i + \theta (x_i - a_i),\ 0 < \theta < 1 \\ &\quad \text{for } i = 1, 2, \cdots r \qquad (7) \end{aligned}$$

where x_i for $i = 1, 2, \cdots r$ will denote here the sample quantities and a_i's the corresponding population values. Each of the sample estimators in the expansion are consistent estimators of the corresponding population parameters and their biases tend to zero with increasing sample sizes. This is because the expected values of each of the squares of the deviations of the estimators from their population parameters tend to zero as the sample sizes tend to infinity. Further the partial derivatives occurring in the expansion are finite and hence, the consistency and asymptotic unbiasedness of Y_h.

Now, the mean square deviation V is greater than or equal to $\phi_h . \sigma_h^2 / n_h''$, as the various correlation and regression coefficients in the estimator Y_h are no longer assumed to be known. In fact these are estimated in a more efficient way than in the above case, where we utilise only the portions of the common units. So the mean square deviation is less than or equal to $\sigma_h^2/n_h'' \; E(\hat{\phi}_h)$ and hence the inequality (5). Further, it can be easily seen that $\hat{\phi}_h$ is an asymptotically unbiased estimator of ϕ_h. So far large sample sizes, the lower and upper limits of V are nearly the same and so a consistent estimator of any one of the limits will be the consistent estimator of V. One such estimator is as given in equation (6).

The proof of $\hat{V}$ being a consistent and asymptotically unbiased estimator of the lower limit of V follows from the fact that $\hat{\phi}_h$ and $\hat{\sigma}_h^2$ are consistent and asymptotically unbiased estimators of the corresponding population parameters and so is $\hat{V}$, a function of $\hat{\phi}_h$ and $\hat{\sigma}_h^2$, of V a function of corresponding population parameters.

4. Loss of Efficiency in Using Modified Weight

In this section we shall prove that the use of modified weight in (1) results in inflating its variance, *i.e.*, loss of efficiency in general, when the various correlation and regression coefficients are estimated in a way such that their distribution is independent of that of the various sample means occurring in (1). One such estimation procedure has been discussed in previous section.

Proof.—Let us first consider an artificial situation where the weights which are functions of correlation coefficients are assumed known constant but the regression coefficients occurring in (1) are estimated from the sample as indicated above. In this case Y_h' is the estimator obtained by minimising the variance of an unbiased linear combination of the two estimators Z_h and $\bar{y}_h''$ and its variance is given by

$$\text{Var}(Y_h') = \phi_h' \frac{\sigma_h^2}{n_h''}. \tag{8}$$

The variance of the other estimator Y_h is given by

$$\begin{aligned}\text{Var}(Y_h) &= (1-\phi_h)^2 \psi + \phi_h^2 \, \text{Var}(\bar{y}_h'') \\ &= \phi_h \left[\frac{(1-\phi_h)}{\theta} + \phi_h \right] \frac{\sigma_h^2}{n_h''}\end{aligned} \tag{9}$$

where θ is a positive proper fraction such that

$$\frac{\phi_h}{1-\phi_h}=\frac{\theta\psi}{\text{Var}\,(\bar{y}_h'')}, \tag{10}$$

ψ being the variance of Z_h under the conditions stated above. Now since $\text{Var}\,(Y_h')\leqslant \text{Var}\,(Y_h)$, so

$$(\theta\phi_h'-\phi_h)\leqslant -\phi_h^2\,(1-\theta). \tag{11}$$

When the weights are also estimated from the sample, the variance of Y_h' is given by

$$\begin{aligned}\text{Var}\,(Y_h') &= E\,[(1-\hat{\phi}_h')^2\,\theta\psi+\hat{\phi}_h'^2\,\text{Var}\,(\bar{y}_h'')]\\ &= \frac{\sigma_h^2}{n_h''}\,E\,[\theta\hat{\phi}_h'+\hat{\phi}_h'^2\,(1-\theta)]\end{aligned} \tag{12}$$

by noting that $\hat{\phi}_h'/(1-\hat{\phi}_h')$ is given by (10) when θ is equated to unity. Also variance of Y_h is given by

$$\text{Var}\,(Y_h)=\frac{\sigma_h^2}{n_h''}\,E\,(\hat{\phi}_h). \tag{13}$$

Now Y_h' the estimator with modified weight has a variance greater than or equal to that of Y_h, the estimator given in (1), if

$$\frac{\sigma_h^2}{n_h''}\,E\,[\theta\,\hat{\phi}_h'-\hat{\phi}_h+\hat{\phi}_h'^2\,(1-\theta)]\geqslant 0$$

which is true if,

$$(\theta\hat{\phi}_h'-\hat{\phi}_h)+\hat{\phi}_h'^2\,(1-\theta)\geqslant 0$$

for every $\hat{\phi}_h$ and $\hat{\phi}_h'$ calculated from a sample. Now the equation (11) is true for every such estimated values of weights. We note therefore that the above inequality is true if

$$(1-\theta)\,(\hat{\phi}_h'^2-\hat{\phi}_h^2)\geqslant 0.$$

Since the weights are positive, this inequality is true if

$$\hat{\phi}_h'\geqslant\hat{\phi}_h$$

or if,

$$\frac{\hat{\phi}_h'}{1-\hat{\phi}_h'}-\frac{\hat{\phi}_h}{1-\hat{\phi}_h}\geqslant 0$$

which is true and so the above statement.

This is to be noted that the modification of weight suggested in (1) by Narain is based on the classical method of weighing two estimators inversely proportional to their variances in order to combine them linearly. This was used earlier by Jessen for $h = 2$ in 1942. By this method here, the weights themselves become function of the parameters to be estimated from the sample. That, this method in such cases can at the most lead to an estimator having minimum variance asymptotically, is clear from the above discussion.

5. Summary

This paper gives the variance of the minimum-variance estimator of the population mean on an occasion in univariate successive sampling, when the various correlation and regression coefficients occurring in the estimator are not known but are estimated from the sample in a specified way. If a more efficient method of estimating consistently various correlation and regression coefficients is used, then the estimator is shown to be a consistent and asymptotically unbiased estimator. In this case the lower and upper limits of the mean square deviation of the estimator is obtained. A consistent and asymptotically unbiased estimator of the variance or mean-square-deviation, depending upon the method of estimation of regression and correlation coefficients, is also given. It is finally shown that the modification, suggested by Narain, in the weight occurring in the estimator given by Patterson and the author results in loss of efficiency of the estimator. This is due to the fact that the classical procedure of weighing two estimators inversely proportional to their variances, where by making the weights themselves functions of the parameters to be estimated from the sample, does not in general lead to the most efficient estimator.

References

1. Narain, R. D. .. "On the Recurrence Formula in Sampling on Successive Occasions," *J. Ind. Soc. Agric. Stat.*, 1953 (1), 96–99.

2. Patterson, H. D. .. "Sampling on Successive Occasions with Partial Replacement of Units," *J.R.S.S.*, 1950, **12 B**, 241–55.

3. Tikkiwal, B. D. .. "Theory of Successive Sampling" (Unpublished *Thesis* submitted towards partial fulfilment of requirements for diploma, I.C.A.R., New Delhi), 1951.