

Some considerations on the ratio and regression estimates

Murakami, Masayasu
Institute of Mathematics, Faculty of Science, Kyushu University

<https://doi.org/10.5109/12950>

出版情報 : 統計数理研究. 4 (1/2), pp.39-42, 1950-12. Research Association of Statistical Sciences

バージョン :

権利関係 :



SOME CONSIDERATIONS ON THE RATIO AND REGRESSION ESTIMATES⁽¹⁾

By M. MURAKAMI

(Institute of Mathematics, Faculty of Science Kyushu University)

1. Introduction

It is well known that in the estimations of population mean and total the ratio and regression estimates are frequently used in the sampling surveys. The theories of these estimates have been greatly developed by many American statisticians. In these theories, however, the cost are not taken into account. The object of this paper is to compare the efficiencies of the ratio estimate and regression estimate with that of mean per s. u. estimate under the cost considerations in the large samples. Let Y be a variate which is the object of estimation in the survey, and X an auxiliary variate correlated with Y . For simplicity, we shall discuss the comparison of efficiencies for these estimates in the simple and double samplings. When the regression method is used, we shall assume that only the variate Y is normally distributed in the simple sampling, and Y and X are normally distributed in the double sampling. But when other method is used we shall not assume the normality of variate. The aim in both methods of estimation is to obtain increased accuracy by taking advantage of the correlation between Y and X . Hereafter the following notations will be used:

- \bar{x}_p : population mean of X
- \bar{y}_p : population mean of Y
- \bar{x}_n : sample mean of X
- \bar{y}_n : sample mean of Y
- ρ : correlation coefficient between X and Y
- C_x : coefficient of variation of X
- C_y : coefficient of variation of Y
- c_1 : cost per unit of X
- c_2 : cost per unit of Y
- C : total cost
- N : total numbers of unit in the population

$$\lambda = \frac{c_1}{c_2}, \quad B = \frac{C_x}{C_y}.$$

2. The case of single sampling

When \bar{x}_p is known, the following Table 1 will be easily obtained according to the current sampling theories.

(1) The results in this paper were communicated at a meeting of the Research Association of Statistical Sciences in Tokyo, December 20, 1949.

TABLE 1. *Costs and variances of three estimates in single sampling.*

method of estimation	estimator	cost	variance (or M. S. E)	variance (or M. S. E) under the cost consideration
mean per S. U.	$Y = \frac{1}{n} \sum_{i=1}^n y_i$	$C = c_2 n$	$\sigma_Y^2 = \frac{N-n}{N} \frac{\sigma_y^2}{n}$	$\sigma_Y^2 = \frac{1}{N} \left(\frac{Nc_2}{C} - 1 \right) \sigma_y^2$
ratio	$Y' = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \bar{x}_p$	$C = (c_1 + c_2)n$	$V(Y') = \frac{N-n}{Nn} \times (1 - \rho B + B^2) \sigma_Y^2$	$V(Y') = \frac{1}{N} \left\{ \frac{(c_1 + c_2)N}{C} - 1 \right\} \times (1 - 2\rho B + B^2) \sigma_y^2$
regression	$Y'' = \bar{y}_n + b(\bar{x}_p - \bar{x}_n)$, where $b = \frac{\sum (x - \bar{x}_n)(y - \bar{y}_n)}{\sum (x - \bar{x}_n)^2}$	$C = (c_1 + c_2)n$	$V(Y'') = \frac{N-n}{Nn} \times (1 - \rho^2) \sigma_y^2$	$V(Y'') = \frac{1}{N} \left\{ \frac{(c_1 + c_2)N}{C} - 1 \right\} \times (1 - \rho^2) \sigma_y^2$

Here we assume that the sample size n is sufficiently large so that approximate formulae for the variances of the ratio and regression estimates are valid.

(1) *Comparison with the ratio estimate and the per s.u.*

It may be observed that $\sigma_Y^2 \cong V(Y')$

according as $\rho \cong \frac{1}{2B} \left\{ \frac{\lambda N}{(1+\lambda)N - C/c_2} + B^2 \right\},$

In particular, as $N \rightarrow \infty$ $\sigma_Y^2 \cong V(Y')$

according as $\rho \cong \frac{1}{2B} \left\{ \frac{\lambda}{1+\lambda} + B^2 \right\}.$

(2) *Comparison with the regression estimate and the per s.u.*

It may be observed that $\sigma_Y^2 \cong V(Y'')$

according as $\rho \cong \sqrt{\frac{\lambda N}{(1+\lambda)N - C/c_2}},$

which is equivalent to, as $N \rightarrow \infty$ $\rho \cong \sqrt{\frac{\lambda}{1+\lambda}}.$

(3) *Comparison with the ratio estimate and the regression estimate*

Since $(1 - 2\rho B + B^2)$ takes the minimum value $1 - \rho^2$ when $B = \rho$, the regression estimate is more accurate than the ratio estimate except when $B = \rho$, in whose case the two variances $V(Y')$, $V(Y'')$ are equal.

3. The case of double sampling

We shall consider an infinite population and take two independent samples of size L and n respectively.

In the first large sample of size L , we shall measure only X and in the second sample of size n both X and Y . For the case of regression estimate

we shall, however, consider moreover another sampling procedure in which the second sample will be drawn from the first sample. The M. S. E. of the estimates for this procedure may be expressed by the notation $V^*(Y'')$.

The current sampling theory will give us the following Table 2. The calculation of $V(Y')$ in the table 2 will be given in Note 1.

TABLE 2. *Costs and variances of three estimate in double sampling*

method of estimation	estimator	cost	variance (or M. S. E)	variance (or M. S. E) under the cost consideration
mean per S. U.	$Y = \frac{1}{n} \sum_1^n y_i$	$C = c_2 n$	$\sigma_{Y'}^2 = \sigma_y^2 / n$	$\sigma_{Y'}^2 = c_2 \sigma_y^2 / C$
ratio	$Y' = \frac{\sum y_i}{\sum x_i} \bar{x}_s$, where $\bar{x}_s = \frac{1}{L} \sum_1^L x_i$	$C = c_1 L$ $+ (c_1 + c_2)n$	$V(Y') = \frac{\sigma_y^2}{n}$ $\times (1 - 2\rho B + B^2)$ $+ \frac{\sigma_y^2 B^2}{L}$	$V(Y') = c_2 \sigma_y^2 / C$ $\times [\sqrt{(1+\lambda)(1-2\rho B+B^2)} + B\sqrt{\lambda}]^2$
regression	$Y'' = \bar{y}_n + b(\bar{x}_s - \bar{x}_n)$, where $b = \frac{\sum (x - \bar{x}_n)(y - \bar{y}_n)}{\sum (x - \bar{x}_n)^2}$	$C = c_1 L$ $+ (c_1 + c_2)n$	$V(Y'') = \sigma_y^2(1 - \rho^2)/n$ $+ \rho^2 \sigma_y^2 / L$	$V(Y'') = c_2 \sigma_y^2 / C$ $\times [\sqrt{(1+\lambda)(1-\rho^2)} + \rho \sqrt{\lambda}]^2$
		$C = c_1 L + c_2 n$	$V^*(Y'') = \sigma_y^2(1 - \rho^2)/n$ $+ \rho^2 \sigma_y^2 / L$	$V^*(Y'') = c_2 \sigma_y^2 / C$ $\times [\sqrt{1-\rho^2} + \rho \sqrt{\lambda}]^2$

(1) *Comparison with σ_y^2 and $V(Y')$*

When $\tilde{x}_p \gg \sigma_x^2 / L$, it follows that $\sigma_{Y'}^2 \cong V(Y')$ according as

$$\rho \cong \frac{\sqrt{\lambda}}{1+\lambda} \left(1 + \frac{\lambda + B^2}{2\sqrt{\lambda} B} \right),$$

which becomes, when $B=1$,

(2) *Comparison with σ_y^2 and $V^*(Y'')$*

It is easily shown that $\sigma_{Y'}^2 \cong V^*(Y'')$ according as

$$|\rho| \cong \frac{2\sqrt{\lambda}}{1+\lambda}.$$

(3) *Comparison with $V(Y')$, $V(Y'')$ and $V^*(Y'')$*

If $B \geq |\rho|$, then $V(Y') \cong V(Y'')$. On the other hand, $V(Y') \cong V^*(Y'')$, however ρ may be.

Appendix.

NOTE 1. The proof of $V(Y')$ may be as below shown;

Let us put
$$Y' = \frac{\sum_1^n y_i}{\sum_1^n x_i} \bar{x}_L = Z \cdot W,$$

where
$$\bar{x}_i = \frac{1}{L} \sum_1^L x_j.$$

Since Z and W are mutually independent it follows that

$$E(Y') = \frac{\tilde{y}_p}{\tilde{x}_p} \cdot \tilde{x}_p = \tilde{y}_p.$$

The second moment of Z about the origin is

$$E(Z^2) = \frac{K}{n} + R_p^2,$$

where
$$K = \frac{\sigma_y^2}{\tilde{x}_p^2} (1 - 2\rho B + B^2)$$

and
$$R_p = \frac{\tilde{y}_p}{\tilde{x}_p}.$$

Hence, the second moment of Y' about the origin is

$$E(Y'^2) = \left(\frac{K}{n} + R_p^2 \right) \left(\tilde{x}_p^2 + \frac{\sigma_x^2}{L} \right).$$

Therefore, we shall find that

$$V(Y') = \frac{K}{n} (\tilde{x}_p^2 + \sigma_x^2/L) + \frac{\sigma_y^2 B^2}{L}.$$

If $\tilde{x}_p \gg \sigma_x^2/L$, then obtain we approximately

$$V(Y') = \frac{\sigma_y^2}{n} (1 - 2\rho B + B^2) + \frac{\sigma_y^2 B^2}{L}.$$

NOTE 2. If the variance form is given by

$$V = \sigma_y^2 (A/n + B/L)$$

and the cost form

$$C = an + bL,$$

where A , B , a , and b are constants, then the optimum variance is easily given by

$$V_{opt} = \frac{\sigma_y^2}{C} (\sqrt{aA} + \sqrt{bB})^2.$$

From this result the variances of three estimates under the cost consideration are as shown in the right hand column of the Table 2.

References

- (1) W. G. COCHRAN: Sample survey techniques. (1948).