A NOTE ON MODIFIED RATIO ESTIMATOR

BY

PADAM SINGH, M.P. JHA & H.V.L. BATHLA

I.A.S.R.I., New Delhi-12

(Received : October, 1977)

1. INTRODUCTION

The ratio method of estimation in sampling is recommended only when the regression of the study variable (Y) on the auxiliary variable (X) is linear and passes through the origin. In practice, the regression of Y on X may not be linear and even if it is linear it need not pass through the origin.

The main purpose of this note is to modify the usual ratio method of estimation when regression of Y on X in the population is of the general form,

$$y=f(x) \qquad \ldots(1.1)$$

where $f(x)$ is some function of x. If $f(x)$ is known completely, the usefulness of taking $f(x)$ as an auxiliary character needs no emphasis. In case the function $f(x)$ is not defined completely but its form is known, the use of utilizing this form for ratio method of estimation has been discussed in this note.

2. THE SUGGESTED PROCEDURE

Suppose that the population under consideration consists of N distinct and identifiable units. Let the information on auxiliary variable (X) be available for all the units of the population. Let x_i, y_i be the values for the i-th unit of the population on the character X and Y respectively.

The suggested procedure consists of the following steps :

(*a*) Select a sample of size n from the population by simple random sampling without replacement and observe the value of the character under study on these units.

(*b*) Determine the functional relationship between y and x using least square technique. Suppose the functional relationship is of the form,

$$y=f(x)$$

The constants of $f(x)$ are obtained so that,

$$E=\sum_{i=1}^{n}\left[\, y_i-f(x_i)\,\right]^2 \qquad \ldots(2.1)$$

is minimum.

(*c*) Define a new auxiliary variable **Z** given by different values of $f(x_i)$ and use this new variable for ratio method of estimation and let the estimator of mean so defined be,

$$\bar{y}_{MR}=\frac{\bar{y}_n}{\bar{z}_n}\,.\,\bar{Z}_N \qquad \ldots(2.2)$$

when the function of $f(x)$ is a polynomial of degree K it is shown below that,

$$\bar{z}_n=\bar{y}_n$$

Following the usual notations,

$$z_i=\hat{y}_i=\hat{\alpha}+\hat{\beta}_1 x_i+\hat{\beta}_2 x_i^2+\ldots\ldots+\hat{\beta}_k x_i^k$$

$$=\bar{y}_n+b_1\,(x_i-\bar{x}_n)+b_2\left(x_i^2-\overline{x_n^2}\right)+\ldots\ldots$$

$$+b_k\left(x_i^k-\overline{x_n^k}\right),$$

where, $$\overline{x_n^r}=\frac{1}{n}\sum_{i=1}^{n} x_i^r$$

and $b_1, b_2\ldots\ldots, b_k$ are the partial regression coefficients calculated from the sample.

Thus,

$$\bar{z}_n=\frac{1}{n}\sum_{i=1}^{n}\hat{y}_i$$

$$=\bar{y}_n$$

In this case $\bar{y}_{MR}$ reduces to $\bar{Z}_N$.

It is easy to see that the large sample variance of the modified ratio estimator is given by,

$$V(\bar{y}_{MR})=V(\bar{y}_n)+R^2_{yz} \,.\, V(\bar{z}_n)-2R_{yz}\,\text{Cov.}\,(\bar{y}_n,\bar{z}_n) \qquad ...(2.3)$$

As the error in estimating the constants of the relationship is small, the above expression simplifies to,

$$V(\bar{y}_{MR})=V(\bar{y}_n)-\text{Cov.}\,(\bar{y}_n,\bar{z}_n),\ \text{as}\ R_{yz}=\frac{\bar{y}_N}{\bar{z}_N}\simeq 1$$

$$=V(\bar{y}_n)\left(1-\eta^2_{yz}\right)$$

$$=\left(\frac{1}{n}-\frac{1}{N}\right)S^2_y\left(1-\eta^2_{yz}\right) \qquad ...(2.4)$$

where η_{yz} is the correlation ratio between the character under study and the auxiliary character.

It may be remarked here that as $n \rightarrow N$ or $\eta_{yz}=1$, the variance of the modified ratio estimator tends to zero. Also, when $\eta_{yz}=0$, the variance reduces to the variance of the simple mean.

3. Efficiency Comparison

The variances of the ratio and regression estimators are known to be

$$V(\bar{y}_R)=\left(\frac{1}{n}-\frac{1}{N}\right)\left(S^2_y+R^2\,S^2_x-2RS_{xy}\right) \qquad ...(3.1)$$

and

$$V(\bar{y}_{lr})=\left(\frac{1}{n}-\frac{1}{N}\right)S^2_y\,(1-\rho^2) \qquad ...(3.2)$$

respectively. It can be seen that the modified ratio estimator is even better than the regression estimator as ($|\eta_{yz}|>|\rho|$) which in turn is known to be always better than the ratio estimator. The gain in efficiency is mainly due to the higher order relationship being determined in y and x. However, if the relationship is only linear, the modified estimator has no gain over the regression estimator. The per cent gain in efficiency over the usual regression estimator can easily be calculated from,

$$\xi=\frac{\eta^2-\rho^2}{1-\rho^2}\times 100$$

4. Numerical Illustration

In order to compare the efficiency of the modified ratio estimator with the ratio estimator, the usual regression estimator and the simple mean, the data on yield (Y), leaf length (X_1) and cane thickness (X_2) collected at the Sugarcane Reseach Station, Punjab Agricultural University, Jullunder were utilised. For populations I and II the auxiliary variables used were X_1 and X_2 respectively. The corresponding new auxiliary variables defined were Z_1 and Z_2. The sampling variances of different estimates for $n=40$, $N=200$,

$$S_y^2=4220,\ \rho_{yx_1}=0.492,\ \rho_{yx_2}=0.791,\ \eta_{yz_1}=0.989 \text{ and}$$

$\eta_{yz_2}=0.932$ are being presented in the following Table 1.

TABLE 1

Variances of estimates of mean

Estimate	*Variances for*	
	Population I	*Population II*
Simple mean	84.4	84.4
Regression estimate	65.7	31.6
Modified ratio estimate	1.72	11.1

It can be seen from the above table 1 that the modified ratio estimator is better than the other estimators including regression estimator. The gain in efficiency is also considerable.