Jour. Ind. Soc. Ag. Statistics
50(2), 1997 : 150–155

Ratio Method of Estimation in the Presence of Measurement Errors

Shalabh
University of Jammu, Jammu-180004
(Received : May, 1996)

SUMMARY

In the present article, the effect of measurement errors on the ratio estimation technique of the population mean is examined.

Key words : Ratio method, Measurement error.

1. *Introduction*

In survey sampling, the properties of estimators based on data originating under various kinds of sampling schemes and various ways of estimation procedure are generally analyzed under the supposition that observations have been recorded without any error. Such a supposition may not be tenable in actual practice and the data may contain observational or measurement errors due to various reasons; see, e.g., Cochran [1] and Sukhatme *et al.* [2].

An important source of measurement errors in survey data is the nature of variables. The nature of the variable arising from the definition may often be such that exact measurements on it are not available. This may happen chiefly due to three reasons. First is that the variable is clearly defined but it is hard to take correct observations at least with the currently available techniques or because of other types of practical difficulties. Consequently, imperfect measurements are obtained such as in the case of monthly expenditure on food items in a household or the level of blood sugar in a human being. Second is that the variable is conceptually well defined but observations can be obtained only on some closely related substitutes known as proxies or surrogates. A simple example is the measurement of economic status of a person or the level of education on which true observations cannot be obtained. One may, for instance, use the number of years of schooling for the level of education, and it may serve as a reasonably good approximation. Third is that the variable is fully comprehensible and well understood but it is not intrinsically defined and properly quantified. However, it is known to be closely associated with a number of factors. Simple examples of such variables are intelligence, specific abilities, utility, aggressiveness, etc. Any theoretical characteristic having at least one of these features is popularly labelled as unobservable or latent variable. True observations on such characteristics are generally unavailable. The reported observations are thus contaminated by measurement errors.

In this article, we consider the estimation of population mean arising from a popular method of estimation, viz., ratio method and analyze its properties in the presence of measurement errors.

2. *Main Results*

Suppose that we are given a set of n paired observations obtained through simple random sampling procedure, on two characteristics X and Y. It is assumed that x_i and y_i for the i^{th} sampling unit are recorded instead of true values X_i and Y_i. The observational or measurement errors are defined as

$$u_i = (y_i - Y_i) \tag{2.1}$$

$$v_i = (x_i - X_i) \tag{2.2}$$

which are assumed to be stochastic with mean 0 but possibly different variances σ_u^2 and σ_v^2.

For the sake of simplicity in exposition, we assume that u_i's and v_i's are uncorrelated although X_i's and Y_i's are correlated. Such a specification can be, however, relaxed at the cost of some algebraic complexity. We also assume that finite population correction can be ignored.

Let the population means of X and Y characteristic be μ_X and μ_Y and population variances σ_X^2 and σ_Y^2. Further, let ρ be the population correlation coefficient between X and Y.

For the estimation of population mean μ_Y, the traditional unbiased estimator is the sample mean $\bar{y}$ but it does not utilize the sample information on X characteristic. One popular way to incorporate it is the ratio method. Assuming that μ_X is known and is different from zero, this method yields the following estimator of μ_Y :

$$t_R = \left(\frac{\bar{y}}{\bar{x}}\right)\mu_X \tag{2.3}$$

where $\bar{x}$ denotes the mean of sample observations on X.

In order to study the efficiency properties of the estimators $\bar{y}$ and t_R in the presence of measurement errors, we first introduce the following notations:

$$C_Y = \frac{\sigma_Y}{\mu_Y}, \quad w_u = n^{-1/2}\,\Sigma u_i, \quad w_Y = n^{-1/2}\Sigma(Y_i - \mu_Y)$$

$$C_X = \frac{\sigma_X}{\mu_X}, \quad w_v = n^{-1/2}\,\Sigma v_i, \quad w_X = n^{-1/2}\Sigma(X_i - \mu_X)$$

It is easy to see that

$$(\bar{y} - \mu_Y) = \frac{1}{n}\Sigma\,[(Y_i - \mu_Y) + u_i] \tag{2.4}$$

$$= n^{-1/2}\,(w_Y + w_u)$$

whence we observe that $\bar{y}$ is unbiased with variance

$$V(\bar{y}) = \frac{\sigma_Y^2}{n}\left(1 + \frac{\sigma_u^2}{\sigma_Y^2}\right) \tag{2.5}$$

Next, we can express

$$(t_R - \mu_Y) = \frac{\mu_Y + \dfrac{w_Y + w_u}{n^{1/2}}}{1 + \dfrac{w_X + w_v}{n^{1/2}\mu_X}} - \mu_Y \tag{2.6}$$

$$= \frac{1}{n^{1/2}}\left[w_Y + w_u - \frac{\mu_Y}{\mu_X}(w_X + w_v)\right]\left[1 + \frac{w_X + w_v}{n^{1/2}\,\mu_X}\right]^{-1}$$

$$= \frac{a}{n^{1/2}} - \frac{ab}{n} + O_p(n^{-3/2})$$

where

$$a = w_y + w_u - \frac{\mu_Y}{\mu_X}(w_X + w_v)$$

$$b = \frac{1}{\mu_X}(w_X + w_v) \tag{2.7}$$

Thus the bias and mean squared error of t_R upto order $O(n^{-1})$ are given by

$$B(t_R) = \frac{E(a)}{n^{1/2}} - \frac{E(ab)}{n} \tag{2.8}$$

$$= \frac{\mu_Y}{n}\left[C_X(C_X - \rho C_Y) + \frac{\sigma_v^2}{\mu_X^2} \right]$$

$$M(t_R) = \frac{E(a^2)}{n} \tag{2.9}$$

$$= \frac{\sigma_Y^2}{n}\left[1 - \frac{C_X}{C_Y}\left(2\rho - \frac{C_X}{C_Y} \right)\right] + \frac{1}{n}\left[\sigma_u^2 + \left(\frac{\mu_Y}{\mu_X} \right)^2 \sigma_v^2 \right]$$

Looking at the expressions presented above, we observe that the measurement errors have no influence at all on the unbiasedness of $\bar{y}$ but for the ratio estimator t_R, only the errors in auxiliary characteristic X affect the bias, at least to the order of our approximation.

Examining the expressions (2.5) and (2.9), we observe that sampling variability in each case increases when measurement errors are present. It is interesting to note that the increase in variability attributable to measurement errors is small in case of $\bar{y}$ when compared with that of t_R.

Next, we find that the estimator t_R is superior to $\bar{y}$ with respect to the criterion of mean squared error upto order $O(n^{-1})$ when

$$\rho > \frac{C_X}{2C_Y}\left(1 + \frac{\sigma_v^2}{\sigma_X^2} \right) \quad \text{if } \mu_X \text{ and } \mu_Y \text{ have same signs} \tag{2.10}$$

$$\rho < -\frac{C_X}{2C_Y}\left(1 + \frac{\sigma_v^2}{\sigma_X^2} \right) \quad \text{if } \mu_X \text{ and } \mu_Y \text{ have opposite signs} \tag{2.11}$$

In a particular case where C_X and C_Y are identical in magnitudes, these conditions reduce to the following:

$$\rho > \frac{1}{2}\left(1 + \frac{\sigma_v^2}{\sigma_X^2} \right) \quad \text{if } \mu_X \text{ and } \mu_Y \text{ have same signs} \tag{2.12}$$

$$\rho < -\frac{1}{2}\left(1 + \frac{\sigma_v^2}{\sigma_X^2}\right) \text{ if } \mu_X \text{ and } \mu_Y \text{ have opposite signs} \quad (2.13)$$

Obviously, both of these conditions will not be satisfied if σ_v^2 exceeds σ_X^2. In other words, if the auxiliary characteristic is so poorly measured that error variance σ_v^2 is larger than σ_X^2, then also t_R is beaten by $\bar{y}$ besides the well known case when the inequalities (2.12) and (2.13) hold true with a reversed sign.

The above observations have an interesting implication. Even in those situations where the ratio estimator is known to have better performance than sample mean in the absence of any measurement errors in X characteristic, cases may arise in which ratio estimator turns out to be poor than sample mean in the presence of any measurement errors. In other words, the measurement errors in X characteristic may alter the preference ordering of $\bar{y}$ and t_R derived under the assumption of absence of measurement errors. Interestingly enough, the measurement errors in Y characteristic have no role to play in this kind of preference ordering.

In practice, the conventional formula for the variance of $\bar{y}$ is (σ_Y^2/n). Comparing with the true variance (2.5), we observe that use of (σ_Y^2/n) will lead to an under-reporting of true standard error. Similar is the case when we examine the expression (2.9) for the mean squared error of ratio estimator to the order of our approximation. It is interesting to note that the under-reporting in case of $\bar{y}$ is by an amount (σ_u^2/n) which is smaller than the corresponding quantity in case of t_R; see the second term in square brackets on the right hand side of (2.9).

The consequence of under-reporting of variability can be clearly appreciated. For instance, it may mislead the practitioner about the precision of the estimate. It may provide shorter but incorrect confidence intervals for the population mean. It may tend to reject the null hypothesis while conducting a test of hypothesis about mean.

It may be remarked that we have restricted our attention in two simple estimators, viz., sample mean and ratio estimator. However, the case of product estimator for the population mean can be easily examined on the same lines. It would be interesting to extend our investigations for other kind of similar estimators arising from different sampling schemes. Generalizing the results when more than one auxiliary characteristic are available will be another direction of future work.

REFERENCES

[1] Cochran, W.G., 1963. *Sampling Techniques*, Second Edition, Wiley Eastern Private Limited, New Delhi.

[2] Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Asok, C., 1984. *Sampling Theory of Surveys with Applications*, Iowa State University Press, Ames.