Int. J. Math. And Appl., 6(1–D)(2018), 847–853

ISSN: 2347-1557

Available Online: http://ijmaa.in/

International Journal of Mathematics And its Applications

Advances in Regression Models used for Business Statistics

Manoj Kumar Srivastava[1,*] and Namita Srivastava[2]

1 Department of Statistics, Institute of Social Sciences, Dr. B.R. Ambedkar University, Agra, Uttar Pradesh, India.

2 Department of Statistics, St. John's College, Agra, Uttar Pradesh, India.

Abstract: The performance of an estimator is generally judged on the basis of relative variance, relative standard error or standard error. These measures are generally unknown since they are the function of the population parameter, thus estimated on the basis of sample information. The expressions of relative bias and relative variance of these estimators are obtained when sampling is done from a finite population and the sample size is large. It is also examined that which of these estimators be adopted as a reasonable criterion for judging the performance of an estimator. The results are verified in case of simple random sampling without replacement.

Keywords: Estimation error, relative bias, relative variance, relative standard error.

1. Introduction

In survey sampling the performance of an estimator of the population parameter is generally judged on the basis of the bias and variance of this estimator. The standard error of the estimator i.e. the square root of the variance is then used to obtain the efficiency of the estimator. The bias and standard error, however, give idea about the absolute error incurred in using the estimator in question. An alternative set of the measures which provide the idea about the relative error are given by the relative bias and relative variance or relative standard error of the estimator. If m is the estimator of the parameter M, then $\frac{m-M}{M}$ is the relative error and the $\frac{E(m)-M}{M}$ and $\frac{\sqrt{V(m)}}{M}$ are called the relative bias and relative standard error of the estimator respectively.

Note that the relative bias, standard error or the relative standard error depend on the population parameter and hence are unknown in practice. In order to have some idea about the magnitude of error, we estimate these quantities from the sample information. Thus, from estimating the efficiency of the estimator, we may employ variance or relative variance, standard error or relative standard error. Which of these should be preferred in practice is a crucial question which we would attempt to answer in the present article when sampling is done from a finite population and the sample size is large.

Let Y be the characteristic in question taking value y_i, $i = 1, 2, \ldots, N$ on the unit with lable i in the population of N units. Let a simple random sample of n units be drawn from this population without replacement. If M is the parameter to the estimated and m is a proposed unbiased estimator, then the relative variance (RV) and the relative standard error (RSE) of the estimator m are given by

$$RV(m) = \theta = \frac{V(m)}{M^2}$$

[*] *E-mail: mkiss87@gmail.com*

$$RSE(m) = \theta^{\frac{1}{2}} = \frac{SE(m)}{M}$$

where $V(m) = E(m - M)^2$, $SE(m) = \{V(m)\}^{\frac{1}{2}}$. The sample estimates of $RV(m) = \theta$ and $RSE(m) = \theta^{\frac{1}{2}}$ are given by

$$\widehat{RV}(m) = \widehat{\theta} = \frac{\widehat{V}(m)}{m^2} \tag{1}$$

$$\widehat{RSE}(m) = \widehat{\theta}^{\frac{1}{2}} = \frac{\widehat{SE}(m)}{m} \tag{2}$$

Note that θ is a measure of efficiency of the estimator m of M and $\hat{\theta}$ is its sample estimate which is not necessarily unbiased. Then the performance of the estimator θ is judged on the basis of relative bias, $RB(\hat{\theta})$ and the relative standard error of $\hat{\theta}, RSE(\hat{\theta})$ which are given by

$$RB(\hat{\theta}) = \frac{E(\hat{\theta}) - \theta}{\theta}$$

$$RSE(\hat{\theta}) = \frac{SE(\hat{\theta})}{\theta}$$

2. Relative Bias of the Estimators

Let Y and Z are two random variables with finite means, $E(Y) < \infty, E(Z) < \infty$ and variance and covariance of Y and Z be of order n^{-r}, $r > 0$, usually r is equal to unity. Thus, these variables converge to their respective means in probability. Let $f(Y, Z)$ be some function for which the Taylor's expansion is valid. Then we have the following result:

Lemma 2.1. *For large n, the relative bias of $f(Y, Z)$ as an estimator of $f_0 = f(E(Y), E(Z))$ is given by*

$$RB[f(Y,Z)] = \frac{1}{2f_0}\left[V(Y)f_Y'' + V(Z)f_Z'' + 2\operatorname{cov}(Y,Z)f_{YZ}''\right] \tag{3}$$

where

$$f_Y'' = \frac{\partial^2 f(Y,Z)}{\partial Y^2}, \quad f_Z'' = \frac{\partial^2 f(Y,Z)}{\partial Z^2}, \quad f_{YZ}'' = \frac{\partial^2 f(Y,Z)}{\partial Z \partial Y}$$

Proof. Expanding the function $f(Y, Z)$ by Taylor's series around $(E(Y), E(Z))$ and retaining the terms up to second order partial derivatives, we have

$$f(Y,Z) = f_0 + (Y - E(Y))f_Y' + (Z - E(Z))f_Z' + \frac{1}{2}\left\{(Y - E(Y))^2 f_Y'' + 2(Y - E(Y))(Z - E(Z))f_{YZ}'' + (Z - E(Z))^2 f_Z''\right\} \tag{4}$$

where f_Y' and f_z' are the first order derivatives of $f(Y, Z)$ with respect to Y and Z respectively, evaluated at $(E(Y), E(Z))$. Taking expectation throughout (4) yields

$$E[f(Y,Z)] = f_0 + \frac{1}{2}\left\{V(Y)f_Y'' + V(Z)f_Z'' + 2\operatorname{cov}(Y,Z)f_{YZ}''\right\} \tag{5}$$

Thus, the relative bias is given by

$$RB[f(Y,Z)] = \frac{E[f(Y,Z)] - f_0}{f_0} = \frac{1}{2f_0}\left[V(Y)f_Y'' + V(Z)f_z'' + 2\operatorname{cov}(Y,Z)f_{YZ}''\right]$$

Hence, the lemma. □

In what follows, we shall prove a result concerning the relative bias of the sample estimate of the relative variance i.e. $\hat{\theta} = \widehat{RV}(m)$.

Theorem 2.2. *If sample size n is large so that the terms of order higher than n could be ignored, then the relative bias of $\widehat{RV}(m)$ is given by*

$$RB(\widehat{RV}(m)) = 3RV(m) - \frac{2\,\mathrm{cov}(m, \widehat{V}(m))}{M.V(m)} \tag{6}$$

where, $E(\widehat{V}(m)) = V(m)$.

Proof. Consider the function f in Lemma 2.1 by

$$f(Y,Z) = \frac{Y}{Z^2}$$
$$Y = \widehat{V}m$$

and $Z = m$. We have $E(Y) = V(m)$ and $E(Z) = M$. Further, we have $f''_Y = \mathbf{0}, f''_Z = \frac{6V(m)}{M^4}, f''_YZ = -\frac{2}{M^3}, \quad f_0 = \frac{V(m)}{M^2}$. Using Lemma 2.1, we get

$$\begin{aligned} RB(\widehat{RV}(m)) &= \frac{M^2}{2V(m)}\left[\frac{6\{V(m)\}^2}{M^4} - \frac{4\,\mathrm{cov}(\widehat{V}(m), m)}{M^2}\right] \\ &= \frac{3V(m)}{M^2} - \frac{2\,\mathrm{cov}(\widehat{V}(m), m)}{M \cdot V(m)} \\ &= 3CV(m) - \frac{2\,\mathrm{cov}(\nabla(m), m)}{M \cdot V(3)} \end{aligned}$$

Hence, the theorem. □

In the following theorem, we calculate the relative bias of $\widehat{RSE}(m)$.

Theorem 2.3. *Under the condition of Theorem 2.1, the relative bias of $\widehat{RSE}(m)$ is given by*

$$RB[\widehat{RSE}(m)] = RV(m) - \frac{RV(\widehat{V}(m))}{8} - \frac{\mathrm{cov}(\widehat{V}(m), m)}{2M.V(m)} \tag{7}$$

Proof. Consider the function f in Lemma 2.1 by

$$f(Y,Z) = \frac{Y^{\frac{1}{2}}}{m} = \widehat{RSE}(m)$$
$$Y = \widehat{V}m$$

and $z = m$. We have

$$4M\left(V(m)^{\frac{3}{2}}\right), \quad f''_Z = \frac{2(V(m))^{\frac{1}{2}}}{M^3}, \quad 2\left(V(m)^{\frac{1}{2}}M^2\right), \quad f_0 = \frac{(V(m))^{\frac{1}{2}}}{M}$$

Using these values in Lemma 2.1 gives

$$\begin{aligned} RB[\widehat{RSE}(m)] &= \frac{M}{2(V(m))^{\frac{1}{2}}}\left[-\frac{1}{4M(V(m))^{\frac{3}{2}}}V(\widehat{V}(m)) + \frac{2(V(m))^{\frac{1}{2}}}{M^3}V(m) - \frac{\mathrm{cov}(\widehat{V}(m), m)}{(\widehat{V}(m))^{\frac{1}{2}}M^2}\right] \\ &= \frac{V(m)}{M^2} - \frac{V(\widehat{V}(m))}{8(V(m))^2} - \frac{\mathrm{cov}(\widehat{V}(m), m)}{2M \cdot V(m)} \\ &= RV(m) - \frac{RV(\widehat{V}(m))}{8} - \frac{\mathrm{cov}(\widehat{V}(m), m)}{2M \cdot V(m)} \end{aligned}$$

Hence, the theorem. □

Theorem 2.4. *For large n, we have*

$$RB[\widehat{SE}(m)] = -\frac{1}{8}RV[\widehat{V}(m)] \tag{8}$$

Proof. Let $g(y)$ be a continuous function of a random variable Y for which the conditions of Taylor's expansion $g(y)$ around $E(Y)$ and retaining terms up to second order derivative, we have

$$g(Y) = g_0 + (Y - E(Y))g' + \frac{1}{2}(Y - E(Y))^2 g'' \tag{9}$$

where,

$$g_0 = g(E(Y)) \quad g' = \left(\frac{\partial}{\partial Y}\right) g(Y)\Big|_{Y=E(Y)}, \quad g'' = \left(\frac{\partial^2}{\partial Y^2}\right) g(y)\Big|_{Y=E(Y)}$$

Taking expectation in (9) throughout, we have

$$E[g(Y)] = g_0 + \frac{1}{2}V(Y)g'' \tag{10}$$

Now, as we assumed earlier that is large so that the second term of (10) could be ignored. Consequently $E[g(Y)] = g_0$. Thus, for large n, $g(Y)$ can be used as an unbiased estimator of g_0. The relative bias of $g(Y)$ as an estimator of g_0 is given by

$$RB[g(Y)] = \frac{E[g(Y)] - g_0}{g_0} = \frac{1}{2g_0}V(Y)g'' \tag{11}$$

Take the function g as $g(Y) = Y^{\frac{1}{2}}$, where $Y = \widehat{V}(m)$, so that $E[\widehat{V}(m)] = V(m)$ $g'' = -\frac{1}{4}(V(m))^{-\frac{3}{2}}$ and $g_0 = (V(m))^{\frac{1}{2}}$. Substituting g_0 and g'' in (11), we obtain

$$RB[\widehat{SE}(m)] = -\frac{V[\widehat{V}(m)]}{8(V(m))^2} = -\frac{1}{8}RV[\widehat{V}(m)]$$

$\widehat{RV}(m)$, $\widehat{RSE}(m)$ and $\widehat{SE}(m)$ give us an idea about the error incurred in using an estimator m of M. We can compare these measures in terms of their relative biases. □

Theorem 2.5. *For symmetric distributions*

$$RB[\widehat{SE}(m)] \leq RB[\widehat{RSE}(m)] \leq RB[\widehat{RV}(m)] \tag{12}$$

Proof. From (6)and (7), we get

$$RB[\widehat{RV}(m)] - RB[\widehat{RSE}(m)] = 2RV(m) + \frac{RV[\widehat{V}(m)]}{8} - \frac{3\,\mathrm{cov}(\widehat{V}(m), m)}{2M.V(m)}$$

and from (7) and (8), we get

$$RB[\widehat{RSE}(m)] - RB[\widehat{SE}(m)] = RV(m) - \frac{\mathrm{cov}(\widehat{V}(m), m)}{2M.V(m)}$$

□

For symmetric distributions $\mathrm{cov}(\widehat{V}(m), m)$ vanishes (Kendall and Stuart 1972) yields the required inequality (12). In view of the expression involved, it is highly unlikely that the inequality (12) are violated even for non-symmetrical populations.

3. The Relative Variance of the Estimators

In order to study the sampling fluctuations in the estimators of measures of sampling error, we consider the relative variances of $\widehat{RV}(m), \widehat{RSE}(m)$ and $\widehat{SE}(m)$. Following theorem plays on important role:

Theorem 3.1. *Ignoring the terms of order higher than n^{-r}, we have*

$$RV[\widehat{RV}(m)] = 4.RV[\widehat{RSE}(m)] \tag{13}$$

Proof. Assume that $g(Y)$ is continuous and is a function for which Taylor's series expansion holds. It is easy to verify

$$V[g(Y)] = V(Y)g'^2 \tag{14}$$

where g' is the first derivative of $g(Y)$ with respect to Y evaluated at $Y = E(Y)$. We have

$$\begin{aligned} RV\left(Y^{\frac{1}{2}}\right) &= V\left(\frac{Y^{\frac{1}{2}}}{E\left(Y^{\frac{1}{2}}\right)}\right) = V\left(\frac{Y^{\frac{1}{2}}}{\left(E(Y) - V\left(Y^{\frac{1}{2}}\right)\right)^{\frac{1}{2}}}\right) \\ &= \frac{\frac{V\left(Y^{\frac{1}{2}}\right)}{E(Y)}}{1 - \frac{V\left(Y^{\frac{1}{2}}\right)}{E(Y)}} \end{aligned} \tag{15}$$

Using (14), it could be easily seen that

$$V\left(Y^{\frac{1}{2}}\right) = \frac{V(Y)}{4E(Y)} \tag{16}$$

Substituting this value of $V\left(Y^{\frac{1}{2}}\right)$ in (15), we obtain

$$RV\left(Y^{\frac{1}{2}}\right) = \frac{RV(Y)}{4 - RV(Y)} \tag{17}$$

$RV(Y)$ will be of order n^{-r} whenever $V(Y)$ is of order n^{-r}. Under the assumption that $\frac{RV(Y)}{4} < 1$, which will normally be true, $RV\left(Y^{\frac{1}{2}}\right)$ in (17) could be approximated by $\frac{RV(Y)}{4}$ This gives

$$RV(Y) = 4RV\left(Y^{\frac{1}{2}}\right) \tag{18}$$

On taking $Y = \widehat{RV}(m)$, we have

$$RV(\widehat{RV}(m)) = 4RV(\widehat{RSE}(m))$$

Hence, the theorem. □

Theorem 3.2. *If we retain terms up to the order n^{-r}, we have*

$$RV[\widehat{RSE}(m)] = RV[\widehat{SE}(m)] + RV(m)\frac{\operatorname{cov}(\widehat{V}(m), m)}{M.V(m)} \tag{19}$$

Proof. From (4) and (5), retaining terms of order up to n^{-r}, we can easily show that

$$RV[f(Y,Z)] = \frac{1}{f_0^2}\left[V(Y)f'_Y + V(Z)f'_Z + 2\operatorname{cov}(Y,Z).f'_Y f'_Z\right] \tag{20}$$

Now, let $f(Y, Z) = \frac{Y^{\frac{1}{2}}}{z}$ where, $Y = \widehat{V}(m), z = m$. It is easy to check

$$f'_Y\Big|_{(V(m),M)} = \frac{1}{2M(V(m))^{\frac{1}{2}}}; \quad f'_Z\Big|_{(V(m),M)} = -\frac{(V(m))^{\frac{1}{2}}}{2M^2}, \quad f_0 = \frac{(V(m))^{\frac{1}{2}}}{M}$$

Substituting these values in (20) and simplifying

$$\begin{aligned} RV[\widehat{RSE}(m)] &= \frac{V(\widehat{V}(m))}{4(V(m))^2} + \frac{V(m)}{M^2} - \frac{\operatorname{cov}(\widehat{V}(m), m)}{MV(m)} \\ &= \frac{1}{4} RV(\bar{V}(m)) + RV(m) - \frac{\operatorname{cov}(\bar{V}(m), m)}{MV(m)} \end{aligned}$$

From (18), we obtain

$$RV[\widehat{RSE}(m)] = RV(\widehat{SE}(m)) + RV(m) - \frac{\operatorname{cov}(\widehat{V}(m), m)}{MV(m)}$$

Hence the theorem. □

Theorem 3.3. *For symmetrical populations, i.e* $\operatorname{cov}(\widehat{V}(m), m) = 0$, *we obtain*

$$RV[\widehat{SE}(m)] < RV[\widehat{RSE}(m)] < RV[\widehat{RV}(m)] \tag{21}$$

Note that the last part of the inequality also holds for asymmetrical population.

4. Illustration

Let $y_1, y_2, \ldots, y_n$ be the observations obtained from a simple random sample drawn from a population consisting of N units. Suppose we are interested in estimating population mean $\bar{y}$. The unbiased estimator of population mean is given by sample mean i.e. $\bar{y}_s$. The relative variance of $\bar{y}_s$ is given by

$$\begin{aligned} RV(\bar{y}_s) &= \frac{N-n}{Nn} \frac{S^2}{\bar{y}^2} \\ &\simeq \frac{1}{n} \frac{S^2}{\bar{y}^2} \end{aligned} \tag{22}$$

If fpc is ignored. Moreover, the variance of $\bar{y}_s$, ignoring fpc, is given by

$$V(\bar{y}_s) = \frac{S^2}{n}$$

The unbiased estimator of $V(\bar{y}_s)$ is obtained as

$$\widehat{V}(\bar{y}_s) = \frac{S^2}{n}$$

where s^2 is the sample variance. This gives

$$V\left(\widehat{V}(\bar{y}_s)\right) = \frac{1}{n^2} V\left(s^2\right) = \frac{S^4}{n^2}\left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right]$$

where $\beta_2 = \frac{\mu_4}{S^4}$. This gives

$$RV\left(\widehat{V}(\bar{y}_s)\right) = \frac{V\left(\frac{s^2}{n}\right)}{\frac{S^4}{n^2}} = \left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right] \tag{23}$$

Further, the covariance between $\bar{y}_s$ and $\frac{s^2}{n}$, if fpc is ignored and N is large, using equation 14 on page 239 in Sukhatme et.al. (1984), is given by

$$\text{cov}\left(\bar{y}_s, \frac{s^2}{n}\right) = \frac{\mu_{30}}{n^2}$$

where

$$\mu_{30} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^3$$

This gives

$$\frac{\text{cov}\left(\bar{y}_s, \frac{s^2}{n}\right)}{\bar{y}V(\bar{y}_s)} = \frac{\mu_{30}}{n\bar{y}S^2} \tag{24}$$

Using (22), (23) and (24) in Theorems 2.1, 2.2, 2.3, we obtain

$$\begin{aligned}
RB\left[\widehat{RV}(\bar{y}_s)\right] &= \frac{3S^2}{n\bar{y}^2} - \frac{\mu_{30}}{n\bar{y}S^2} \\
RB\left[\widehat{RSE}(\bar{y}_s)\right] &= \frac{S^2}{ny^2} - \frac{1}{8}\left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right] - \frac{\mu_{30}}{2n\bar{y}S^2} \quad \text{and} \\
RB\left[\widehat{SE}(\bar{y}_s)\right] &= -\frac{1}{8}\left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right]
\end{aligned}$$

For symmetrical population μ_{30} will be zero, we can easily verify the inequality given in Theorem 2.4. Similarly, on using the results in (22), (23), and (24) in Theorems 3.1, 3.2 we get

$$\begin{aligned}
RV\left[\widehat{RSE}(\bar{y}_s)\right] &= \frac{S^2}{n\bar{y}^2} - \frac{1}{4}\left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right] - \frac{\mu_{30}}{n\bar{y}S^2} \\
RV\left[\widehat{RV}(\bar{y}_s)\right] &= 4RV\left[\widehat{RSE}(\bar{y}_s)\right] \quad \text{and} \\
RV\left[\widehat{SE}(\bar{y}_s)\right] &= \frac{1}{4}\left[\frac{\beta_2 - 1}{n} + \frac{2}{n(n-1)}\right]
\end{aligned}$$

Thus, for symmetrical populations, $\mu_{30} = 0$, we can easily verify the inequality show in the Theorem 3.3.

References

[1] W. G. Cochran, *Sampling techniques*, John Wiley & Sons, New York, (1977).

[2] G. S. Bieler and R. L. Williams, *Generalised standard error models for proportions in complex design surveys*, Proceedings of Section on Survey Methods of American Statistical Association, (1990), 272-277.

[3] M. E. Gonzalez, J. L. Ogus, O. G. Shapiro and B. J. Tepping, *Standards for discussion and presentation of errors in survey and census data*, Journal of American Statistical Association, 70(351)(1975), 5-23.

[4] Lepkovski, *Presentation of Sampling Errors*, Methods of Survey Sampling/Applied Sampling-Lecture at Statistics South Africa, (1998).

[5] The Australian Bureau of Statistics, *Technical Note on Sampling Variability*, in ABS-HES Summary of Results, Appendix D, (1993-94), 43-49.

[6] R. Swanepoel and D. J. Sloker, *The estimation and presentation of standard errors in a survey report*, Statistics South Africa, (2000).

[7] P. V. Sukhatme, B. V. Sukhatme, S. Sukhatme and C. Ashok, *Sampling Theory of Surveys with Applications*, Indian Society of Agricultural Statistics, (1984).