Introduction to Sampling Theory

Lecture 16
Ratio Method of Estimation

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

http://home.iitk.ac.in/~shalab/sp

Bias and Mean Squared Error of Ratio Estimator:

Ratio estimator of $\overline{Y}$

$$\hat{\overline{Y}}_R = \frac{\overline{y}}{\overline{x}} \overline{X} = \hat{R}\overline{X}$$

We approximate the bias and mean squared error (mse) as follows:

Let

$$\varepsilon_0 = \frac{\overline{y} - \overline{Y}}{\overline{Y}} \Rightarrow \overline{y} = (1 + \varepsilon_o)\overline{Y}$$

$$\varepsilon_1 = \frac{\overline{x} - \overline{X}}{\overline{X}} \Rightarrow \overline{x} = (1 + \varepsilon_1)\overline{X}.$$

SRSWOR is being followed to draw the sample.

Bias and Mean Squared Error of Ratio Estimator:

$$E(\varepsilon_0) = 0$$

$$E(\varepsilon_1) = 0$$

$$E(\varepsilon_0^2) = \frac{f}{n} C_Y^2$$

$$E(\varepsilon_1^2) = \frac{f}{n} C_X^2$$

$$E(\varepsilon_0 \varepsilon_1) = \frac{1}{\bar{X}\bar{Y}} \frac{f}{n} S_{XY} = \frac{1}{\bar{X}\bar{Y}} \frac{f}{n} \rho S_X S_Y = \frac{f}{n} \rho \frac{S_X}{\bar{X}} \frac{S_Y}{\bar{Y}} = \frac{f}{n} \rho C_X C_Y$$

where $f = \frac{N-n}{N}, \; S_Y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$

$C_X = \frac{S_X}{\bar{X}}, C_Y = \frac{S_Y}{\bar{Y}}$**: coefficient of variation related to** Y.

Bias and Mean Squared Error of Ratio Estimator:

Writing $\hat{\bar{Y}}_R$ in terms of ε's, we get

$$\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X} = \frac{(1+\varepsilon_0)\bar{Y}}{(1+\varepsilon_1)\bar{X}}\bar{X}$$

$$= (1+\varepsilon_0)(1+\varepsilon_1)^{-1}\bar{Y}.$$

Assuming $|\varepsilon_1| < 1$, the term $(1+\varepsilon_1)^{-1}$ may be expanded as an infinite series and it would be convergent.

Bias and Mean Squared Error of Ratio Estimator:

Assuming $|\varepsilon_1| < 1$ **means that**

$$\left|\frac{\bar{x} - \bar{X}}{\bar{X}}\right| < 1,$$

i.e., possible estimate $\bar{X}$ **of population mean** $\bar{X}$ **lies between 0 and** $2\bar{X}$.

This is likely to hold true if the variation in $\bar{X}$ **is not large.**

In order to ensures that variation in $\bar{X}$ **is small, assume that the sample size** n **it is fairly large.**

Bias and Mean Squared Error of Ratio Estimator:

With this assumption

$$\hat{\bar{Y}}_R = \bar{Y}(1+\varepsilon_0)(1-\varepsilon_1+\varepsilon_1^2-\ldots)$$

$$= \bar{Y}(1+\varepsilon_0-\varepsilon_1+\varepsilon_1^2-\varepsilon_1\varepsilon_0+\ldots).$$

So the estimation error of $\hat{\bar{Y}}_R$ **is**

$$\hat{\bar{Y}}_R - \bar{Y} = \bar{Y}(\varepsilon_0-\varepsilon_1+\varepsilon_1^2-\varepsilon_1\varepsilon_0+\ldots).$$

Bias and Mean Squared Error of Ratio Estimator:

In case, when sample size is large, then ε_0 and ε_1 are likely to be small quantities and so the terms involving second and higher powers of ε_0 and ε_1 would be negligibly small.

In such a case,

$$\hat{\bar{Y}}_R - \bar{Y} \simeq \bar{Y}(\varepsilon_0 - \varepsilon_1)$$

and

$$E(\hat{\bar{Y}}_R - \bar{Y}) = 0.$$

So the ratio estimator is an unbiased estimator of population mean up to the first order of approximation.

Bias and Mean Squared Error of Ratio Estimator:

If we assume that only terms of ε_0 and ε_1 involving powers more than two are negligibly small (which is more realistic than assuming that powers more than one are negligibly small), then the estimation error of $\hat{\bar{Y}}_R$ can be approximated as

$$\hat{\bar{Y}}_R - \bar{Y} \simeq \bar{Y}(\varepsilon_0 - \varepsilon_1 + \varepsilon_1^2 - \varepsilon_1\varepsilon_0).$$

Bias and Mean Squared Error of Ratio Estimator:

Then the bias of $\hat{\bar{Y}}_R$ is given by

$$E(\hat{\bar{Y}}_R - \bar{Y}) = \bar{Y}\left(0 - 0 + \frac{f}{n}C_X^2 - \frac{f}{n}\rho C_X C_y\right)$$

$$Bias(\hat{\bar{Y}}_R) = E(\hat{\bar{Y}}_R - \bar{Y}) = \frac{f}{n}\bar{Y}C_X(C_X - \rho C_Y)$$

up to the second order of approximation.

The bias generally decreases as the sample size grows large.

Bias and Mean Squared Error of Ratio Estimator:

The bias of $\hat{\bar{Y}}_R$ is zero, i.e.,

$$Bias(\hat{\bar{Y}}_R) = 0$$

if $$E(\varepsilon_1^2 - \varepsilon_0\varepsilon_1) = 0$$

or if $$\frac{Var(\bar{x})}{\bar{X}^2} - \frac{Cov(\bar{x},\bar{y})}{\bar{X}\bar{Y}} = 0$$

or if $$\frac{1}{\bar{X}^2}\left[Var(\bar{x}) - \frac{\bar{X}}{\bar{Y}}Cov(\bar{x},\bar{y})\right] = 0$$

or if $$Var(\bar{x}) - \frac{Cov(\bar{x},\bar{y})}{R} = 0 \quad (\textbf{assuming } \bar{X} \neq 0)$$

or if $$R = \frac{\bar{Y}}{\bar{X}} = \frac{Cov(\bar{x},\bar{y})}{Var(\bar{x})}$$

which is satisfied when the regression line of *Y* on *X* passes through origin.

Bias and Mean Squared Error of Ratio Estimator:

Now, to find the mean squared error, consider

$$MSE(\hat{\bar{Y}}_R) = E(\hat{\bar{Y}}_R - \bar{Y})^2$$

$$= E\left[\bar{Y}^2(\varepsilon_0 - \varepsilon_1 + \varepsilon_1^2 - \varepsilon_1\varepsilon_0 + \ldots)^2\right].$$

Under the assumption $|\varepsilon_1| < 1$ **and the terms of** ε_0 **and** ε_1 **involving powers more than two are negligible small,**

$$MSE(\hat{\bar{Y}}_R) \simeq E\left[\bar{Y}^2(\varepsilon_0^2 + \varepsilon_1^2 - 2\varepsilon_0\varepsilon_1)\right]$$

$$= \bar{Y}^2\left[\frac{f}{n}C_X^2 + \frac{f}{n}C_Y^2 - \frac{2f}{n}\rho C_X C_Y\right]$$

$$= \frac{\bar{Y}^2 f}{n}\left[C_X^2 + C_Y^2 - 2\rho C_X C_y\right]$$

up to the second order of approximation.