Chapter 5

Ratio and Product Methods of Estimation

An important objective in any statistical estimation procedure is to obtain the estimators of parameters of interest with more precision. It is also well understood that incorporation of more information in the estimation procedure yields better estimators, provided the information is valid and proper. Use of such auxiliary information is made through the ratio method of estimation to obtain an improved estimator of the population mean. In ratio method of estimation, auxiliary information on a variable is available, which is linearly related to the variable under study and is utilized to estimate the population mean.

Let Y be the variable under study and X be an auxiliary variable which is correlated with Y. The observations x_i on X and y_i on Y are obtained for each sampling unit. The population mean $\bar{X}$ of X (or equivalently the population total X_{tot}) must be known. For example, $x_i's$ may be the values of $y_i's$ from

- some earlier completed census,
- some earlier surveys,
- some characteristic on which it is easy to obtain information etc.

For example, if y_i is the quantity of fruits produced in the i^{th} plot, then x_i can be the area of i^{th} plot or the production of fruit in the same plot in the previous year.

Let $(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)$ be the random sample of size n on the paired variable (X, Y) drawn, preferably by SRSWOR, from a population of size N. The ratio estimate of the population mean $\bar{Y}$ is

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

assuming the population mean $\bar{X}$ is known. The ratio estimator of population total $Y_{tot} = \sum_{i=1}^{N} Y_i$ is

$$\hat{Y}_{R(tot)} = \frac{y_{tot}}{x_{tot}} X_{tot}$$

where $X_{tot} = \sum_{i=1}^{N} X_i$ is the population total of X which is assumed to be known, $y_{tot} = \sum_{i=1}^{n} y_i$ and $x_{tot} = \sum_{i=1}^{n} x_i$ are the sample totals of Y and X respectively. The $\hat{Y}_{R(tot)}$ can be equivalently expressed as

$$\begin{aligned}\hat{Y}_{R(tot)} &= \frac{\bar{y}}{\bar{x}} X_{tot} \\ &= \hat{R} X_{tot}.\end{aligned}$$

Looking at the structure of ratio estimators, note that the ratio method estimates the relative change $\frac{Y_{tot}}{X_{tot}}$ that occurred after (x_i, y_i) were observed. It is clear that if the variation among the values of $\frac{y_i}{x_i}$ and is nearly same for all i = 1,2,...,n then values of $\frac{y_{tot}}{x_{tot}}$ (or equivalently $\frac{\bar{y}}{\bar{x}}$) vary little from sample to sample and the ratio estimate will be of high precision.

Bias and mean squared error of ratio estimator:

Assume that the random sample $(x_i, y_i), i = 1, 2, ..., n$ is drawn by SRSWOR and population mean $\bar{X}$ is known. Then

$$E(\hat{\bar{Y}}_R) = \frac{1}{\binom{N}{n}} \sum_{i=1}^{\binom{N}{n}} \frac{\bar{y}_i}{\bar{x}_i} \bar{X}$$
$$\neq \bar{Y} \text{ (in general).}$$

Moreover, it is difficult to find the exact expression for $E\left(\frac{\bar{y}}{\bar{x}}\right)$ and $E\left(\frac{\bar{y}^2}{\bar{x}^2}\right)$. So we approximate them and proceed as follows:
Let

$$\varepsilon_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}} \Rightarrow \bar{y} = (1 + \varepsilon_o)\bar{Y}$$
$$\varepsilon_1 = \frac{\bar{x} - \bar{X}}{\bar{X}} \Rightarrow \bar{x} = (1 + \varepsilon_1)\bar{X}.$$

Since SRSWOR is being followed, so

$$E(\varepsilon_0) = 0$$
$$E(\varepsilon_1) = 0$$
$$E(\varepsilon_0^2) = \frac{1}{\bar{Y}^2} E(\bar{y} - \bar{Y})^2$$
$$= \frac{1}{\bar{Y}^2} \frac{N-n}{Nn} S_Y^2$$
$$= \frac{f}{n} \frac{S_Y^2}{\bar{Y}^2}$$
$$= \frac{f}{n} C_Y^2$$

where $f = \frac{N-n}{N}$, $S_Y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ and $C_Y = \frac{S_Y}{\bar{Y}}$ is the coefficient of variation related to Y.

Similarly,

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

$$E(\varepsilon_0 \varepsilon_1) = \frac{1}{\bar{X}\bar{Y}} E[(\bar{x} - \bar{X})(\bar{y} - \bar{Y})]$$

$$= \frac{1}{\bar{X}\bar{Y}} \frac{N-n}{Nn} \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y})$$

$$= \frac{1}{\bar{X}\bar{Y}} . \frac{f}{n} S_{XY}$$

$$= \frac{1}{\bar{X}\bar{Y}} \frac{f}{n} \rho S_X S_Y$$

$$= \frac{f}{n} \rho \frac{S_X}{\bar{X}} \frac{S_Y}{\bar{Y}}$$

$$= \frac{f}{n} \rho C_X C_Y$$

where $C_X = \frac{S_X}{\bar{X}}$ is the coefficient of variation related to X and ρ is the population correlation coefficient between X and Y.

Writing $\hat{\bar{Y}}_R$ in terms of $\varepsilon's,$ 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. Such an assumption means that $\left| \frac{\bar{x} - \bar{X}}{\bar{X}} \right| < 1,$ i.e., a possible estimate $\bar{x}$ of the population mean $\bar{X}$ lies between 0 and 2 $\bar{X}$. This is likely to hold if the variation in $\bar{x}$ is not large. In order to ensure that variation in $\bar{x}$ is small, assume that the sample size n is fairly large. With this assumption,

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

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

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

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

In case, when the 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 the population mean up to the first order of approximation.

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

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

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

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

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

upto the second order of approximation. The bias generally decreases as the sample size grows large.

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 $Var(\bar{x}) - \frac{Cov(\bar{x},\bar{y})}{R} = 0$ (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 the origin.

Now, to find the mean squared error, consider

$$\begin{aligned} MSE(\hat{\bar{Y}}_R) &= E(\hat{\bar{Y}}_R - \bar{Y})^2 \\ &= E\left[\bar{Y}^2(\varepsilon_0 - \varepsilon_1 + \varepsilon_1^2 - \varepsilon_1\varepsilon_0 + \ldots)^2\right] \\ &\simeq E\left[\bar{Y}^2(\varepsilon_0^2 + \varepsilon_1^2 - 2\varepsilon_0\varepsilon_1)\right]. \end{aligned}$$

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

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

up to the second-order of approximation.

Efficiency of ratio estimator in comparison to SRSWOR

Ratio estimator is a better estimate of $\bar{Y}$ than sample mean based on SRSWOR if

$$MSE(\hat{\bar{Y}}_R) < Var_{SRS}(\bar{y})$$

or if $\bar{Y}^2\frac{f}{n}(C_X^2 + C_Y^2 - 2\rho C_X C_Y) < \bar{Y}^2\frac{f}{n}C_Y^2$

or if $C_X^2 - 2\rho C_X C_Y < 0$

or if $\rho > \frac{1}{2}\frac{C_X}{C_Y}$.

Thus ratio estimator is more efficient than the sample mean based on SRSWOR if

$$\rho > \frac{1}{2}\frac{C_X}{C_Y} \quad \text{if} \quad R>0$$

and $\rho < -\frac{1}{2}\frac{C_X}{C_Y}$ if $R<0$.

It is clear from this expression that the success of ratio estimator depends on how close is the auxiliary information to the variable under study.

Upper limit of ratio estimator:

Consider

$$Cov(\hat{R},\bar{x}) = E(\hat{R}\,\bar{x}) - E(\hat{R})E(\bar{x})$$
$$= E\left(\frac{\bar{y}}{\bar{x}}\bar{x}\right) - E(\hat{R})E(\bar{x})$$
$$= \bar{Y} - E(\hat{R})\bar{X}.$$

Thus

$$E(\hat{R}) = \frac{\bar{Y}}{\bar{X}} - \frac{Cov(\hat{R},\bar{x})}{\bar{X}}$$
$$= R - \frac{Cov(\hat{R},\bar{x})}{\bar{X}}$$

$$Bias(\hat{R}) = E(\hat{R}) - R$$
$$= -\frac{Cov(\hat{R}, \bar{x})}{\bar{X}}$$
$$= -\frac{\rho_{\hat{R},\bar{x}}\sigma_{\hat{R}}\sigma_{\bar{x}}}{\bar{X}}$$

where $\rho_{\hat{R},\bar{x}}$ is the correlation between $\hat{R}$ and $\bar{x}$; $\sigma_{\hat{R}}$ and $\sigma_{\bar{x}}$ are the standard errors of $\hat{R}$ and $\bar{x}$ respectively.

Thus

$$\left|Bias(\hat{R})\right| = \frac{\left|-\rho_{\hat{R},\bar{x}}\right|\sigma_{\hat{R}}\sigma_{\bar{x}}}{\bar{X}}$$
$$\leq \frac{\sigma_{\hat{R}}\sigma_{\bar{x}}}{\bar{X}} \quad \left(\left|\rho_{\hat{R},\bar{x}}\right| \leq 1\right).$$

assuming $\bar{X} > 0$. Thus

$$\left|\frac{Bias(\hat{R})}{\sigma_{\hat{R}}}\right| \leq \frac{\sigma_{\bar{x}}}{\bar{X}}$$

or $$\left|\frac{Bias(\hat{R})}{\sigma_{\hat{R}}}\right| \leq C_X$$

where C_X is the coefficient of variation of X. If $C_X < 0.1$, then the bias in $\hat{R}$ may be safely regarded as negligible in relation to the standard error of $\hat{R}$.

Alternative form of $MSE(\hat{Y}_R)$

Consider

$$\sum_{i=1}^{N}(Y_i - RX_i)^2 = \sum_{i=1}^{N}\left[(Y_i - \bar{Y}) + (\bar{Y} - RX_i)\right]^2$$
$$= \sum_{i=1}^{N}\left[(Y_i - \bar{Y}) - R(X_i - \bar{X})\right]^2 \quad (\text{Using } \bar{Y} = R\bar{X})$$
$$= \sum_{i=1}^{N}(Y_i - \bar{Y})^2 + R^2\sum_{i=1}^{N}(X_i - \bar{X})^2 - 2R\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})$$

$$\frac{1}{N-1}\sum_{i=1}^{N}(Y_i - RX_i)^2 = S_Y^2 + R^2S_X^2 - 2RS_{XY}.$$

The *MSE* of $\hat{\bar{Y}}_R$ has already been derived which is now expressed again as follows:

$$MSE(\hat{\bar{Y}}_R) = \frac{f}{n}\bar{Y}^2(C_Y^2 + C_X^2 - 2\rho C_X C_Y)$$
$$= \frac{f}{n}\bar{Y}^2\left(\frac{S_Y^2}{\bar{Y}^2} + \frac{S_X^2}{\bar{X}^2} - 2\frac{S_{XY}}{\bar{X}\bar{Y}}\right)$$
$$= \frac{f}{n}\frac{\bar{Y}^2}{\bar{Y}^2}\left(S_Y^2 + \frac{\bar{Y}^2}{\bar{X}^2}S_X^2 - 2\frac{\bar{Y}}{\bar{X}}S_{XY}\right)$$
$$= \frac{f}{n}\left(S_Y^2 + R^2S_X^2 - 2RS_{XY}\right)$$
$$= \frac{f}{n(N-1)}\sum_{i=1}^{N}(Y_i - RX_i)^2$$
$$= \frac{N-n}{nN(N-1)}\sum_{i=1}^{N}(Y_i - RX_i)^2.$$

Estimate of $MSE(\hat{Y}_R)$

Let $U_i = Y_i - RX_i$, $i = 1,2,..,N$ then MSE of $\hat{\bar{Y}}_R$ can be expressed as

$$MSE(\hat{\bar{Y}}_R) = \frac{f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(U_i - \bar{U})^2 = \frac{f}{n}S_U^2$$

where $S_U^2 = \frac{1}{N-1}\sum_{i=1}^{N}(U_i - \bar{U})^2.$

Based on this, a natural estimator of $MSE(\hat{\bar{Y}}_R)$ is

$$MSE(\hat{\bar{Y}}_R) = \frac{f}{n}s_u^2$$

where $s_u^2 = \frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2$

$$= \frac{1}{n-1}\sum_{i=1}^{n}\left[(y_i - \bar{y}) - \hat{R}(x_i - \bar{x})\right]^2$$
$$= s_y^2 + \hat{R}^2s_x^2 - 2\hat{R}s_{xy},$$

$\hat{R} = \frac{\bar{y}}{\bar{x}}.$

Based on the expression

$$MSE(\hat{\bar{Y}}_R) = \frac{f}{n(N-1)}\sum_{i=1}^{N}(Y_i - RX_i)^2,$$

an estimate of $MSE(\hat{\bar{Y}}_R)$ is

$$MSE(\hat{\bar{Y}}_R) = \frac{f}{n(n-1)}\sum_{i=1}^{n}(y_i - \hat{R}x_i)^2$$
$$= \frac{f}{n}(s_y^2 + \hat{R}^2s_x^2 - 2\hat{R}s_{xy}).$$

Confidence interval of ratio estimator

If the sample is large so that the normal approximation is applicable, then the 100(1-α)% confidence intervals of $\bar{Y}$ and *R* are

$$\left(\hat{\bar{Y}}_R - Z_{\frac{\alpha}{2}}\sqrt{Var(\hat{\bar{Y}}_R)},\quad \hat{\bar{Y}}_R + Z_{\frac{\alpha}{2}}\sqrt{Var(\hat{\bar{Y}}_R)}\right)$$

and

$$\left(\hat{R} - Z_{\frac{\alpha}{2}}\sqrt{Var(\hat{R})},\quad \hat{R} + Z_{\frac{\alpha}{2}}\sqrt{Var(\hat{R})}\right)$$

respectively where $Z_{\frac{\alpha}{2}}$ is the normal derivate to be chosen for a given value of confidence coefficient $(1-\alpha)$.

If $(\bar{x},\bar{y})$ follows a bivariate normal distribution, then $(\bar{y}-R\bar{x})$ is normally distributed. If SRS is followed for drawing the sample, then assuming *R* is known, the statistic

$$\frac{\bar{y}-R\bar{x}}{\sqrt{\frac{N-n}{Nn}(s_y^2+R^2s_x^2-2Rs_{xy})}}$$

is approximately $N(0,1)$.

This can also be used for finding confidence limits, see Cochran (1977, Chapter 6, page 156) for more details.

Conditions under which the ratio estimate is optimum

The ratio estimate $\hat{\bar{Y}}_R$ is the best linear unbiased estimator of $\bar{Y}$ when

(i) the relationship between y_i and x_i is linear passing through origin., i.e.

$$y_i = \beta x_i + e_i,$$

where $e_i's$ are independent with $E(e_i / x_i) = 0$ and β is the slope parameter

(ii) this line is proportional to x_i, i.e.

$$Var(y_i / x_i) = E(e_i^2) = Cx_i$$

where *C* is constant.

Proof. Consider the linear estimate of β because $\hat{\beta} = \sum_{i=1}^{n} \ell_i y_i$ where $y_i = \beta x_i + e_i$ and ℓ_i 's are constant.

Then $\hat{\beta}$ is unbiased if $\bar{Y} = \beta \bar{X}$ as $E(y) = \beta \bar{X} + E(e_i / x_i)$.

If *n* sample values of x_i are kept fixed and then in repeated sampling

$$E(\hat{\beta}) = \sum_{i=1}^{n} \ell_i x_i \beta$$

$$\text{and } Var(\hat{\beta}) = \sum_{i=1}^{n} \ell_i^2 Var(y_i / x_i) = C \sum_{i=1}^{n} \ell_i^2 x_i$$

So $E(\hat{\beta}) = \beta$ when $\sum_{i=1}^{n} \ell_i x_i = 1.$

Consider the minimization of $Var(y_i / x_i)$ subject to the condition for being the unbiased estimator $\sum_{i=1}^{n} \ell_i x_i = 1$ using the Lagrangian function. Thus the Lagrangian function with Lagrangian multiplier is

$$\varphi = Var(y_i / x_i) - 2\lambda(\sum_{i=1}^{n} \ell_i x_i - 1.)$$

$$= C \sum_{i=1}^{n} \ell_1^2 x_i - 2\lambda(\sum_{i=1}^{n} \ell_i x_i - 1).$$

Now

$$\frac{\partial \varphi}{\partial \ell_i} = 0 \Rightarrow \ell_i x_i = \lambda x_i,\ i = 1, 2, .., n$$

$$\frac{\partial \varphi}{\partial \lambda} = 0 \Rightarrow \sum_{i=1}^{n} \ell_i x_i = 1$$

Using $\sum_{i=1}^{n} \ell_i x_i = 1$

or $\sum_{i=1}^{n} \lambda x_i = 1$

or $\lambda = \frac{1}{n\bar{x}}.$

Thus

$$\ell_i = \frac{1}{n\bar{x}}$$

and so $\hat{\beta} = \frac{\sum_{i=1}^{n} y_i}{n\bar{x}} = \frac{\bar{y}}{\bar{x}}.$

Thus $\hat{\beta}$ is not only superior to $\bar{y}$ but also the best in the class of linear and unbiased estimators.

Alternative approach:

This result can alternatively be derived as follows:

The ratio estimator $\hat{R} = \frac{\bar{y}}{\bar{x}}$ is the best linear unbiased estimator of $R = \frac{\bar{Y}}{\bar{X}}$ if the following two conditions hold:

(i) For fixed x, $E(y) = \beta x$, i.e., the line of regression of y on x is a straight line passing through the origin.

(ii) For fixed x, $Var(x) \propto x$, i.e., $Var(x) = \lambda x$ where λ is constant of proportionality.

Proof: Let $\underline{y} = (y_{1)}, y_2, ..., y_n)'$ and $\underline{x} = (x_1, x_2, ..., x_n)'$ be two vectors of observations on $y's$ and $x's$. Hence for any fixed $\underline{x}$,

$$E(\underline{y}) = \beta \underline{x}$$
$$Var(\underline{y}) = \Omega = \lambda \, \text{diag}(x_1, x_2, ..., x_n)$$

where $\text{diag}(x_1, x_2, ..., x_n)$ is the diagonal matrix with $x_1, x_2, ..., x_n$ as the diagonal elements.

The best linear unbiased estimator of β is obtained by minimizing

$$S^2 = (\underline{y} - \beta \underline{x})' \Omega^{-1} (\underline{y} - \beta \underline{x})$$
$$= \sum_{i=1}^{n} \frac{(y_i - \beta x_i)^2}{\lambda x_i}.$$

Solving

$$\frac{\partial S^2}{\partial \beta} = 0$$
$$\Rightarrow \sum_{i=1}^{n} (y_i - \hat{\beta} x_i) = 0$$

or $\hat{\beta} = \frac{\bar{y}}{\bar{x}} = \hat{R}$.

Thus $\hat{R}$ is the best linear unbiased estimator of R. Consequently, $\hat{R}\bar{X} = \hat{\bar{Y}}_R$ is the best linear unbiased estimator of $\bar{Y}$.

Ratio estimator in stratified sampling

Suppose a population of size *N* is divided into *k* strata. The objective is to estimate the population mean $\bar{Y}$ using the ratio method of estimation.

In such a situation, a random sample of size n_i is being drawn from the i^{th} strata of size N_i on the variable under study *Y* and auxiliary variable *X* using SRSWOR.

Let

y_{ij} : j^{th} observation on *Y* from i^{th} strata

x_{ij} : j^{th} observation on *X* from i^{th} strata $i = 1, 2, \ldots, k$; $j = 1, 2, ..., n_i$.

An estimator of $\bar{Y}$ based on the philosophy of stratified sampling can be derived in the following two possible ways:

1. Separate ratio estimator

- Employ first the ratio method of estimation separately in each stratum and obtain ratio estimator $\hat{\bar{Y}}_{R_i}$ $i = 1, 2, .., k$, assuming the stratum mean $\bar{X}_i$ to be known.
- Then combine all the estimates using weighted arithmetic mean.

This gives the separate ratio estimator as

$$\hat{\bar{Y}}_{Rs} = \sum_{i=1}^{k} \frac{N_i \hat{\bar{Y}}_{R_i}}{N}$$

$$= \sum_{i=1}^{k} w_i \hat{\bar{Y}}_{R_i}$$

$$= \sum_{i=1}^{k} w_i \frac{\bar{y}_i}{\bar{x}_i} \bar{X}_i$$

where $\bar{y}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$: sample mean of *Y* from i^{th} strata

$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$: sample mean of *X* from i^{th} strata

$\bar{X}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{ij}$: mean of all the *X* units in i^{th} stratum

No assumption is made that the true ratio remains constant from stratum to stratum. It depends on information on each $\bar{X}_i$.

2. Combined ratio estimator:

- Find first the stratum mean of $Y's$ and $X's$ as

$$\bar{y}_{st} = \sum_{i=1}^{k} w_i \bar{y}_i$$

$$\bar{x}_{st} = \sum_{i=1}^{k} w_i \bar{x}_i.$$

- Then define the combined ratio estimator as

$$\hat{\bar{Y}}_{Rc} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X}$$

where $\bar{X}$ is the population mean of X based on all the $N = \sum_{i=1}^{N} N_i$ units. It does not depend on individual stratum units. It does not depend on information on each $\bar{X}_i$ but only on $\bar{X}$.

Properties of separate ratio estimator:

Note that there is an analogy between $\bar{Y} = \sum_{i=1}^{k} w_i \bar{Y}_i$ and $\bar{Y}_{Rs} = \sum_{i=1}^{k} w_i \bar{Y}_{Ri}$.

We already have derived the approximate bias of $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}} \bar{X}$ as

$$E(\hat{\bar{Y}}_R) = \bar{Y} + \frac{\bar{Y}f}{n}(C_x^2 - \rho C_X C_Y).$$

So for $\hat{\bar{Y}}_{Ri}$, we can write

$$E(\hat{\bar{Y}}_{Ri}) = \bar{Y}_i + \bar{Y}_i \frac{f_i}{n_i}(C_{ix}^2 - \rho_i C_{iX} C_{iY})$$

where $\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$, $\bar{X}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} x_{ij}$

$$f_i = \frac{N_i - n_i}{N_i}, C_{iy}^2 = \frac{S_{iy}^2}{\bar{Y}_i^2}, \; C_{ix}^2 = \frac{S_{ix}^2}{\bar{X}_i^2},$$

$$S_{iy}^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(Y_{ij} - \bar{Y}_i)^2, \; S_{ix}^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(X_{ij} - \bar{X}_i)^2,$$

ρ_i : correlation coefficient between the observation on X and Y in i^{th} stratum

C_{ix} : coefficient of variation of X values in i^{th} sample.

Thus

$$\begin{aligned} E(\hat{\bar{Y}}_{Rs}) &= \sum_{i=1}^{k} w_i E(\hat{\bar{Y}}_{Ri}) \\ &= \sum_{i=1}^{k} w_i \left[\bar{Y}_i + \bar{Y}_i \frac{f_i}{n_i}(C_{ix}^2 - \rho_i C_{ix} C_{iy} \right] \\ &= \bar{Y} + \sum_{i=1}^{k} \frac{w_i \bar{Y}_i f_i}{n_i}(C_{ix}^2 - \rho_i C_{ix} C_{iy}) \end{aligned}$$

$$Bias(\hat{\bar{Y}}_{Rs}) = E(\hat{\bar{Y}}_{Rs}) - \bar{Y}$$
$$= \sum_{i=1}^{k} \frac{w_i \bar{Y}_i f_i}{n_i} C_{ix}(C_{ix} - \rho_i C_{iy})$$

upto the second order of approximation.

Assuming finite population correction to be approximately 1, $n_i = n/k$ and C_{ix}, C_{iy} and ρ_i are the same for all the strata as C_x, C_y and ρ respectively, we have

$$Bias(\hat{\bar{Y}}_{Rs}) = \frac{k}{n}(C_x^2 - \rho C_x C_y).$$

Thus the bias is negligible when the sample size within each stratum should be sufficiently large and $\bar{Y}_{Rs}$ is unbiased when $C_{ix} = \rho C_{iy}$.

Now we derive the approximate *MSE* of $\hat{\bar{Y}}_{Rs}$. We already have derived the *MSE* of $\hat{\bar{Y}}_R$ earlier as

$$MSE(\hat{\bar{Y}}_R) = \frac{\bar{Y}^2 f}{n}(C_X^2 - C_Y^2 - 2\rho C_x C_y)$$
$$= \frac{f}{n(N-1)} \sum_{i=1}^{N} (Y_i - RX_i)^2$$

where $R = \frac{\bar{Y}}{\bar{X}}$.

Thus the MSE of ratio estimate up to the second order of approximation based on i^{th} stratum is

$$MSE(\hat{\bar{Y}}_{Ri}) = \frac{f_i}{n_i(N_i - 1)}(C_{iX}^2 - C_{iY}^2 - 2\rho_i C_{iX} C_{iY})$$
$$= \frac{f_i}{n_i(N_i - 1)} \sum_{j=1}^{N_i} (Y_{ij} - R_i X_{ij})^2$$

and so

$$MSE(\hat{\bar{Y}}_{Rs}) = \sum_{i=1}^{k} w_i^2 \, MSE(\hat{\bar{Y}}_{Ri})$$
$$= \sum_{i=1}^{k} \left[\frac{w_i^2 f_i}{n_i} \bar{Y}_i^2 (C_{iX}^2 + C_{iY}^2 - 2\rho_i C_{iX} C_{iY}) \right]$$
$$= \sum_{i=1}^{k} \left[w_i^2 \frac{f_i}{n_i(N_i - 1)} \sum_{j=1}^{N_i} (Y_{ij} - R_i X_{ij})^2 \right]$$

An estimate of $MSE(\hat{\bar{Y}}_{Rs})$ can be found by substituting the unbiased estimators of S_{iX}^2, S_{iY}^2 and S_{iXY}^2 as s_{ix}^2, s_{iy}^2 and s_{ixy}, respectively for i^{th} stratum and $R_i = \bar{Y}_i / \bar{X}_i$ can be estimated by $r_i = \bar{y}_i / \bar{x}_i$.

$$MSE(\hat{\bar{Y}}_{Rs}) = \sum_{i=1}^{k} \left[\frac{w_i^2 f_i}{n_i}(s_{iy}^2 + r_i^2 s_{ix}^2 - 2 r_i s_{ixy}) \right].$$

Also

$$MSE(\hat{\bar{Y}}_{Rs}) = \sum_{i=1}^{k} \left[\frac{w_i^2 f_i}{n_i(n_i - 1)} \sum_{j=1}^{n_i} (y_{ij} - r_i x_{ij})^2 \right]$$

Properties of combined ratio estimator:

Here

$$\hat{\bar{Y}}_{RC} = \frac{\sum_{i=1}^{k} w_i \bar{y}_i}{\sum_{i=1}^{k} w_i \bar{x}_i} \bar{X} = \frac{\bar{y}_{st}}{\bar{x}_{st}} \bar{X} = \hat{R}_c \bar{X}.$$

It is difficult to find the exact expression of bias and mean squared error of $\hat{\bar{Y}}_{Rc}$, so we find their approximate expressions.

Define

$$\varepsilon_1 = \frac{\bar{y}_{st} - \bar{Y}}{\bar{Y}}$$

$$\varepsilon_2 = \frac{\bar{x}_{st} - \bar{X}}{\bar{X}}$$

$$E(\varepsilon_1) = 0$$

$$E(\varepsilon_2) = 0$$

$$E(\varepsilon_1^2) = \sum_{i=1}^{k} \frac{N_i - n_i}{N_i n_i} \frac{w_i^2 S_{iY}^2}{\bar{Y}^2} = \sum_{i=1}^{k} \frac{f_i}{n_i} \frac{w_i^2 S_{iY}^2}{\bar{Y}^2} \quad \left(\text{Recall that in case of } \hat{\bar{Y}}_R, \ E(\varepsilon_1^2) = \frac{f}{n} \frac{S_Y^2}{\bar{Y}^2} = \frac{f}{n} C_Y^2 \right)$$

$$E(\varepsilon_2^2) = \sum_{i=1}^{k} \frac{f_i}{n_i} \frac{w_i^2 S_{iX}^2}{\bar{X}^2}$$

$$E(\varepsilon_1 \varepsilon_2) = \sum_{i=1}^{k} w_i^2 \frac{f_i}{n_i} \frac{S_{iXY}}{\bar{X}\bar{Y}}.$$

Thus assuming $|\varepsilon_2| < 1$,

$$\begin{aligned} \hat{\bar{Y}}_{RC} &= \frac{(1+\varepsilon_1)\bar{Y}}{(1+\varepsilon_2)\bar{X}} \bar{X} \\ &= \bar{Y}(1+\varepsilon_1)(1-\varepsilon_2+\varepsilon_2^2-\ldots) \\ &= \bar{Y}(1+\varepsilon_1-\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_2^2-\ldots) \end{aligned}$$

Retaining the terms up to order two due to the same reason as in the case of $\hat{\bar{Y}}_R$,

$$\hat{\bar{Y}}_{RC} \simeq \bar{Y}(1+\varepsilon_1-\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_2^2)$$

$$\hat{\bar{Y}}_{RC} - \bar{Y} = \bar{Y}(\varepsilon_1-\varepsilon_2-\varepsilon_1\varepsilon_2+\varepsilon_2^2)$$

The approximate bias of $\hat{\bar{Y}}_{Rc}$ up to the second-order of approximation is

$$
\begin{aligned}
Bias(\hat{\bar{Y}}_{Rc}) &= E(\hat{\bar{Y}}_{Rc} - \bar{Y}) \\
&\simeq \bar{Y}E(\varepsilon_1 - \varepsilon_2 - \varepsilon_1\varepsilon_2 + \varepsilon_2^2) \\
&= \bar{Y}\left[0 - 0 - E\left(\varepsilon_1\varepsilon_2\right) + E\left(\varepsilon_2^2\right)\right] \\
&= \bar{Y}\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2\left(\frac{S_{iX}^2}{\bar{X}^2} - \frac{S_{iXY}}{\bar{X}\bar{Y}}\right)\right] \\
&= \bar{Y}\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2\left(\frac{S_{iX}^2}{\bar{X}^2} - \frac{\rho_i S_{iX}S_{iY}}{\bar{X}\bar{Y}}\right)\right] \\
&= \frac{\bar{Y}}{\bar{X}}\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2 S_{iX}\left(\frac{S_{iX}}{\bar{X}} - \frac{\rho_i S_{iY}}{\bar{Y}}\right)\right] \\
&= R\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2 S_{iX}\left(C_{iX} - \rho_i C_{iY}\right)\right]
\end{aligned}
$$

where $R = \frac{\bar{Y}}{\bar{X}}$, ρ_i is the correlation coefficient between the observations on Y and X in the i^{th} stratum, C_{ix} and C_{iy} are the coefficients of variation of X and Y respectively in the ith stratum.

The mean squared error upto second order of approximation is

$$
\begin{aligned}
MSE(\hat{\bar{Y}}_{Rc}) &= E(\hat{\bar{Y}}_{Rc} - \bar{Y})^2 \\
&\simeq \bar{Y}^2E(\varepsilon_1 - \varepsilon_2 - \varepsilon_1\varepsilon_2 + \varepsilon_2)^2 \\
&\simeq \bar{Y}^2E(\varepsilon_1^2 + \varepsilon_2^2 - 2\varepsilon_1\varepsilon_2) \\
&= \bar{Y}^2\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2\left(\frac{S_{iX}^2}{\bar{X}^2} + \frac{S_{iY}^2}{\bar{Y}^2} - \frac{2S_{iXY}}{\bar{X}\bar{Y}}\right)\right] \\
&= \bar{Y}^2\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2\left(\frac{S_{iX}^2}{\bar{X}^2} + \frac{S_{iY}^2}{\bar{Y}^2} - \frac{2\rho_i S_{iX}}{\bar{X}}\frac{S_{iY}}{\bar{Y}}\right)\right] \\
&= \frac{\bar{Y}^2}{\bar{Y}^2}\sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2\left(\frac{\bar{Y}^2}{\bar{X}^2}S_{iX}^2 + S_{iY}^2 - 2\rho_i\frac{\bar{Y}}{\bar{X}}S_{iX}S_{iY}\right)\right] \\
&= \sum_{i=1}^{k}\left[\frac{f_i}{n_i}w_i^2(R^2S_{iX}^2 + S_{iY}^2 - 2\rho_i RS_{iX}S_{iY})\right].
\end{aligned}
$$

An estimate of $MSE(\bar{Y}_{Rc})$ can be obtained by replacing S_{iX}^2, S_{iY}^2 and S_{iXY} by their unbiased estimators s_{ix}^2, s_{iy}^2 and s_{ixy} respectively whereas $R = \frac{\bar{Y}}{\bar{X}}$ is replaced by $r = \frac{\bar{y}}{\bar{x}}$. Thus the following estimate is obtained:

$$
MSE(\bar{Y}_{Rc}) = \sum_{i=1}^{k}\left[\frac{w_i^2 f_i}{n_i}\left(r^2 s_{ix}^2 + s_{iy}^2 - 2rs_{ixy}\right)\right].
$$

Comparison of combined and separate ratio estimators

An obvious question arises that which of the estimates $\hat{\bar{Y}}_{Rs}$ or $\hat{\bar{Y}}_{Rc}$ is better. So we compare their *MSE*s. Note that the only difference in the term of these *MSE*s is due to the form of ratio estimate. It is

- $R_i = \frac{\bar{y}_i}{\bar{x}_i}$ in $MSE(\hat{\bar{Y}}_{Rs})$
- $\bar{R} = \frac{\bar{Y}}{\bar{X}}$ in $MSE(\hat{\bar{Y}}_{Rc})$.

Thus

$$\begin{aligned}\Delta &= MSE(\hat{\bar{Y}}_{Rc}) - MSE(\hat{\bar{Y}}_{Rs}) \\ &= \sum_{i=1}^{k}\left[\frac{w_i^2 f_i}{n_i}\left[(R^2 - R_i^2)S_{iX}^2 + 2(R_i - R)\rho_i S_{iX} S_{iY}\right]\right] \\ &= \sum_{i=1}^{k}\left[\frac{w_i^2 f_i}{n_i}\left[(R - R_i)^2 S_{iX}^2 + 2(R - R_i)(R_i S_{iX}^2 - \rho_i S_{iX} S_{iY})\right]\right].\end{aligned}$$

The difference Δ depends on

(i) The magnitude of the difference between the strata ratios (R_i) and whole population ratio (*R).*

(ii) The value of $(R_i S_{ix}^2 - \rho_i S_{ix} S_{iy})$ is usually small and vanishes when the regression line of *y* on *x* is linear and passes through origin within each stratum. See as follows:

$$R_i S_{ix}^2 - \rho_i S_{ix} S_{iy} = 0$$

$$R_i = \frac{\rho_i S_{ix} S_{iy}}{S_{ix}^2}$$

which is the estimator of the slope parameter in the regression of *y* on *x* in the i^{th} stratum. In such a case

$$MSE(\hat{\bar{Y}}_{Rc}) > MSE(\hat{\bar{Y}}_{Rs})$$

but $$Bias(\hat{\bar{Y}}_{Rc}) < Bias(\hat{\bar{Y}}_{Rs}).$$

So unless R_i varies considerably, the use of $\hat{\bar{Y}}_{Rc}$ would provide an estimate of $\bar{Y}$ with negligible bias and precision as good as $\hat{\bar{Y}}_{Rs}$.

- If $R_i \neq R$, $\hat{\bar{Y}}_{Rs}$ can be more precise but bias may be large.
- If $R_i \simeq R$, $\hat{\bar{Y}}_{Rc}$ can be as precise as $\hat{\bar{Y}}_{Rs}$ but its bias will be small. It also does not require knowledge of $\bar{X}_1, \bar{X}_2, ..., \bar{X}_k$.

Ratio estimators with reduced bias:

The ratio type estimators that are unbiased or have smaller bias than $\hat{R}, \hat{\bar{Y}}_R$ or $\hat{Y}_{Rc(tot)}$ are useful in sample surveys. There are several approaches to derive such estimators. We consider here two such approaches:

1. Unbiased ratio – type estimators:

Under SRS, the ratio estimator has form $\frac{\bar{Y}}{\bar{x}}\bar{X}$ to estimate the population mean $\bar{Y}$. As an alternative to this, we consider following as an estimator of the population mean

$$\hat{\bar{Y}}_{Ro} = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{X_i}\right)\bar{X} .$$

Let $R_i = \frac{Y_i}{X_i},\ i = 1,2,..,N,$

then

$$\begin{aligned}\hat{\bar{Y}}_{R0} &= \frac{1}{n}\sum_{i=1}^{n} R_i \bar{X} \\ &= \bar{r}\bar{X}\end{aligned}$$

where

$$\bar{r} = \frac{1}{n}\sum_{i=1}^{n} R_i$$

$$\begin{aligned}Bias\ (\hat{\bar{Y}}_{R0}) &= E(\hat{\bar{Y}}_{R0}) - \bar{Y} \\ &= E(\bar{r}\bar{X}) - \bar{Y} \\ &= E(\bar{r})\bar{X} - \bar{Y}.\end{aligned}$$

Since

$$\begin{aligned}E(\bar{r}) &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{N}\sum_{i=1}^{N} R_i\right) \\ &= \frac{1}{n}\sum_{i=1}^{n}\bar{R} \\ &= \bar{R}.\end{aligned}$$

So $Bias(\hat{\bar{Y}}_{R0}) = \bar{R}\bar{X} - \bar{Y}.$

Using the result that under SRSWOR, $Cov(\bar{x},\bar{y}) = \frac{N-n}{Nn} S_{XY}$, it also follows that

$$\begin{aligned}
Cov(\bar{r},\bar{x}) &= \frac{N-n}{Nn}\frac{1}{N-1}\sum_{i=1}^{N}(R_i-\bar{R})(X_i-\bar{X}) \\
&= \frac{N-n}{Nn}\frac{1}{N-1}\left(\sum_{i=1}^{N} R_i X_i - N\bar{R}\bar{X}\right) \\
&= \frac{N-n}{n}\frac{1}{N-1}\left(\sum_{i=1}^{N}\frac{Y_i}{X_i} X_i - N\bar{R}\bar{X}\right) \\
&= \frac{N-n}{Nn}\frac{1}{N-1}(N\bar{Y} - N\bar{R}\bar{X}) \\
&= \frac{N-n}{n}\frac{1}{N-1}[-Bias(\hat{\bar{Y}}_{R0})].
\end{aligned}$$

Thus using the result that in SRSWOR, $Cov(\bar{x},\bar{y}) = \frac{N-n}{Nn} S_{XY}$, and therefore $Cov(\bar{r},\bar{x}) = \frac{N-n}{Nn} S_{RX}$, we have

$$\begin{aligned}
Bias(\hat{\bar{Y}}_{Ro}) &= -\frac{n(N-1)}{N-n} Cov(\bar{r},\bar{x}) \\
&= -\frac{n(N-1)}{N-n}\frac{N-n}{Nn} S_{RX} \\
&= -\left(\frac{N-1}{N}\right) S_{RX}
\end{aligned}$$

where $S_{RX} = \frac{1}{N-1}\sum_{i=1}^{N}(R_i-\bar{R})(X_i-\bar{X}).$

The following result helps in obtaining an unbiased estimator of a population mean:

Since under SRSWOR set up,

$$E(s_{xy}) = S_{xy}$$

where $s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}),$

$$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y}).$$

So an unbiased estimator of the bias in $Bias(\hat{\bar{Y}}_{R0}) = -(N-1)S_{RX}$ is obtained as follows:

$$\begin{aligned}
Bias(\hat{\bar{Y}}_{R0}) &= -\frac{(N-1)}{N} s_{rx} \\
&= -\frac{N-1}{N(n-1)}\sum_{i=1}^{n}(r_i-\bar{r})(x_i-\bar{x}) \\
&= -\frac{N-1}{N(n-1)}\left(\sum_{i=1}^{n} r_i x_i - n\bar{r}\,\bar{x}\right) \\
&= -\frac{N-1}{N(n-1)}\left(\sum_{i=1}^{n}\frac{y_i}{x_i} x_i - n\bar{r}\,\bar{x}\right) \\
&= -\frac{N-1}{N(n-1)}(n\bar{y} - n\bar{r}\,\bar{x}).
\end{aligned}$$

So

$$Bias(\hat{\bar{Y}}_{R0}) = E\left(\hat{\bar{Y}}_{R0}\right) - \bar{Y} = -\frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\,\bar{x}).$$

Thus

$$E\left[\hat{\bar{Y}}_{R0} - Bias(\hat{\bar{Y}}_{R0})\right] = \bar{Y}$$

or $$E\left[\hat{\bar{Y}}_{R0} + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\,\bar{x})\right] = \bar{Y}.$$

Thus

$$\hat{\bar{Y}}_{R0} + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\,\bar{x}) = \bar{r}\bar{X} + \frac{n(N-1)}{N(n-1)}(\bar{y} - \bar{r}\,\bar{x})$$

is an unbiased estimator of the population mean.

It may noted that such an unbiased estimator cannot be obtained using $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X}$ because it is not exactly unbiased. It is unbiased only up to the first order of approximation. So even if the bias of $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X}$ upto the first order of approximation is used to obtain such an unbiased estimator, the estimator will change for higher order of approximations.

2. Jackknife method for obtaining a ratio estimate with lower bias

Jackknife method is used to get rid of the term of order 1/*n* from the bias of an estimator. Suppose the $E(\hat{R})$ can be expanded after ignoring finite population correction as

$$E(\hat{R}) = R + \frac{a_1}{n} + \frac{a_2}{n^2} + \ldots$$

Let *n* = *mg* and the sample is divided at random into *g* groups, each of size *m*. Then

$$E(g\hat{R}) = gR + \frac{ga_1}{gm} + \frac{ga_2}{g^2m^2} + \ldots$$

$$= gR + \frac{a_1}{m} + \frac{a_2}{gm^2} + \ldots$$

Let $\hat{R}_i^* = \frac{\sum^* y_i}{\sum^* x_i}$ where the $\sum^*$ denotes the summation over all values of the sample except the i^{th} group.

So $\hat{R}_i^*$ is based on a simple random sample of size $m(g-1)$, so we can express

$$E(\hat{R}_i^*) = R + \frac{a_1}{m(g-1)} + \frac{a_2}{m^2(g-1)^2} + \ldots$$

or

$$E\left[(g-1)\hat{R}_i^*\right] = (g-1)R + \frac{a_1}{m} + \frac{a_2}{m^2(g-1)} + \ldots$$

Thus

$$E\left[g\hat{R} - (g-1)\hat{R}_i^*\right] = R - \frac{a_2}{g(g-1)m^2} + \ldots$$

or

$$E\left[g\hat{R} - (g-1)\hat{R}_i^*\right] = R - \frac{a_2}{n^2}\frac{g}{g-1} + \ldots$$

Hence the bias of $\left[g\hat{R} - (g-1)\hat{R}_i^*\right]$ is of order $\frac{1}{n^2}$.

Now *g* estimates of this form can be obtained, one estimator for each group. Then the jackknife or Quenouille's estimator is the average of these of estimators

$$\hat{R}_Q = g\hat{R} - (g-1)\frac{\sum_{i=1}^{g}\hat{R}_i}{g}.$$

Product method of estimation:

The ratio estimator is more efficient than the sample mean under SRSWOR if $\rho > \frac{1}{2}.\frac{C_x}{C_y}$, if $R > 0$, which is usually the case. This shows that if auxiliary information is such that $\rho < -\frac{1}{2}\frac{C_x}{C_y}$, then we cannot use the ratio method of estimation to improve the sample mean as an estimator of the population mean. So there is a need for another type of estimator which also makes use of information on auxiliary variable *X*. Product estimator is an attempt in this direction.

The product estimator of the population mean $\bar{Y}$ is defined as

$$\hat{\bar{Y}}_P = \frac{\bar{y}\,\bar{x}}{\bar{X}}.$$

assuming the population mean $\bar{X}$ to be known

We now derive the bias and variance of $\hat{\bar{Y}}_p$.

Let $\varepsilon_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$, $\varepsilon_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}$,

(i) Bias of $\hat{\bar{Y}}_p$.

We write $\hat{\bar{Y}}_p$ as

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

Taking expectation, we obtain bias of $\hat{\bar{Y}}_p$ as

$$Bias(\hat{\bar{Y}}_p) = E(\varepsilon_0\varepsilon_1)\frac{1}{\bar{X}}Cov(\bar{y},\bar{x}) = \frac{f}{n\bar{X}}S_{xy},$$

which shows that bias of $\hat{\bar{Y}}_p$ decreases as n increases. Bias of $\hat{\bar{Y}}_p$ can be estimated by

$$Bias(\hat{\bar{Y}}_p) = \frac{f}{n\bar{X}}s_{xy}.$$

(ii) MSE of $\hat{\bar{Y}}_p$:

Writing $\hat{\bar{Y}}_p$ is terms of ε_0 and ε_1, we find that the mean squared error of the product estimator $\hat{\bar{Y}}_p$ up to second order of approximation is given by

$$MSE(\hat{\bar{Y}}_p) = E(\hat{\bar{Y}}_p - \bar{Y})^2$$
$$= \bar{Y}^2E(\varepsilon_1+\varepsilon_0+\varepsilon_1\varepsilon_2)^2$$
$$\approx \bar{Y}^2E(\varepsilon_1^2+\varepsilon_0^2+2\varepsilon_1\varepsilon_2).$$

Here terms in $(\varepsilon_1,\varepsilon_0)$ of degrees greater than two are assumed to be negligible. Using the expected values, we find that

$$MSE(\hat{\bar{Y}}_p) = \frac{f}{n}\left[S_Y^2 + R^2S_X^2 + 2RS_{XY}\right].$$

(iii) Estimation of MSE of $\hat{\bar{Y}}_p$

The mean squared error of $\hat{\bar{Y}}_p$ can be estimated by

$$MSE(\hat{\bar{Y}}_p) = \frac{f}{n}\left[s_y^2 + r^2s_x^2 + 2rs_{xy}\right]$$

where $r = \bar{y}/\bar{x}$.

(iv) Comparison with SRSWOR:

From the variances of the sample mean under SRSWOR and the product estimator, we obtain

$$Var(\bar{y})_{SRS} - MSE(\hat{\bar{Y}}_p) = -\frac{f}{n} RS_X (2\rho S_Y + RS_X),$$

where $Var(\bar{y})_{SRS} = \frac{f}{n} S_Y^2$ which shows that $\hat{\bar{Y}}_p$ is more efficient than the simple mean $\bar{y}$ for

$$\rho < -\frac{1}{2}\frac{C_x}{C_y} \text{ if } R>0$$

and for

$$\rho > -\frac{1}{2}\frac{C_x}{C_y} \text{ if } R<0.$$

Multivariate Ratio Estimator

Let y be the study variable and $X_1, X_2, ..., X_p$ be p auxiliary variables assumed to be corrected with y. Further, it is assumed that $X_1, X_2, ..., X_p$ are independent. Let $\bar{Y}, \bar{X}_1, \bar{X}_2, ..., \bar{X}_p$ be the population means of the variables y, $X_1, X_2, ..., X_p$. We assume that a SRSWOR of size n is selected from the population of N units. The following notations will be used.

S_i^2 : the population mean sum of squares for the variate X_i ,

s_i^2 : the sample mean sum of squares for the variate X_i,

S_0^2 : the population mean sum of squares for the study variable y,

s_0^2 : the sample mean sum of squares for the study variable y,

$C_i = \frac{S_i}{X_i}$: coefficient of variation of the variate X_i,

$C_0 = \frac{S_0}{Y}$: coefficient of variation of the variate y,

$\rho_i = \frac{S_{iy}}{S_i S_0}$: coefficient of correlation between y and X_i,

$\hat{\bar{Y}}_{Ri} = \frac{\bar{y}}{\bar{X}_i}$: ratio estimator of $\bar{Y}$, based on X_i

where $i = 1, 2, ..., p.$ Then the multivariate ratio estimator of $\bar{Y}$ is given as follows.

$$\hat{\bar{Y}}_{MR} = \sum_{i=1}^{p} w_i \hat{\bar{Y}}_{Ri}, \quad \sum_{i=1}^{p} w_i = 1$$

$$= \bar{y} \sum_{i=1}^{p} w_i \frac{\bar{X}_i}{\bar{x}_i}.$$

(i) Bias of the multivariate ratio estimator:

The approximate bias of $\hat{\bar{Y}}_{Ri}$ up to the second order of approximation is

$$Bias(\hat{\bar{Y}}_{Ri}) = \frac{f}{n}\bar{Y}(C_i^2 - \rho_i C_i C_0).$$

The bias of $\hat{\bar{Y}}_{MR}$ is obtained as

$$Bias(\hat{\bar{Y}}_{MR}) = \sum_{i=1}^{p} w_i \frac{\bar{Y}f}{n}(C_i^2 - \rho_i C_i C_0)$$
$$= \frac{\bar{Y}f}{n}\sum_{i=1}^{p} w_i C_i (C_i - \rho_i C_0).$$

(ii) Variance of the multivariate ratio estimator:

The variance of $\hat{\bar{Y}}_{Ri}$ up to the second-order of approximation is given by

$$Var(\hat{\bar{Y}}_{Ri}) = \frac{f}{n}\bar{Y}^2(C_0^2 + C_i^2 - 2\rho_i C_0 C_i).$$

The variance of $\hat{\bar{Y}}_{MR}$ up to the second-order of approximation is obtained as

$$Var(\hat{\bar{Y}}_{MR}) = \frac{f}{n}\bar{Y}^2\sum_{i=1}^{p} w_i^2(C_0^2 + C_i^2 - 2\rho_i C_0 C_i).$$