# CHAPTER 8

# Ratio Method of Estimation

## 8.1 INTRODUCTION

In Chapter 5, it has been shown that when the auxiliary information has a good relation with the study variable, one can use the auxiliary information to construct various sampling deigns such as probability proportional to size with replacement (PPSWR), probability proportional to size without replacement (PPSWOR), inclusion probability proportion to size (IPPS), and Rao-Hartley-Cochran (RHC) to yield efficient estimators for population parameters (e.g., total, mean, and proportions) under some specific circumstances. Furthermore, we have also mentioned that the auxiliary information can be used gainfully in constructing strata for stratified sampling. The auxiliary information can also be used at the stages of estimation to improve the efficiency of the conventional estimator. In this section we consider the ratio method of estimation to improve the efficiency of the conventional estimator by utilizing auxiliary information $x$, whose values need not be known for all units, but the total $X$ is assumed to be known. For instance, in estimating the total production of a crop, one can use the area under cultivation ($x$) of the farm growing a particular crop as a size variable because production is strictly related to the area under cultivation. Here, the values of the area under cultivation for all the farms need not be known but the total area under cultivation for the crop ($X$) should be known.

## 8.2 RATIO ESTIMATOR FOR POPULATION RATIO

We are very often interested in estimating the ratio $R$ of two characters such as per capita income, proportion of unemployed persons, average cost of treatment per indoor patients, or average cost of educating a person. The ratio estimator of the ratio $R = Y/X$ of the population totals $Y$ and $X$ of the characters $y$ and $x$ is defined as

$$\widehat{R} = \widehat{Y}\Big/\widehat{X} \tag{8.2.1}$$

where $\widehat{X}$ and $\widehat{Y}$ are unbiased estimators of the population totals $X$ and $Y$, respectively.

*Survey Sampling Theory and Applications*
ISBN 978-0-12-811848-1
http://dx.doi.org/10.1016/B978-0-12-811848-1.00008-X

### 8.2.1 Exact Expression of Bias and Mean-Square Error

The estimator $\widehat{R}$ is not unbiased for $R$ in general since $E\left(\widehat{R}\right) \neq E\left(\widehat{Y}\right)/E\left(\widehat{X}\right)$. The bias of $\widehat{R}$ is

$$\begin{aligned} B\left(\widehat{R}\right) &= E\left(\widehat{R}\right) - R \\ &= -\frac{1}{X}\left\{Y - XE\left(\widehat{R}\right)\right\} \\ &= -\frac{1}{X}\left\{E\left(\widehat{R}\,\widehat{X}\right) - E\left(\widehat{X}\right)E\left(\widehat{R}\right)\right\} \\ &= -\frac{1}{X}Cov\left(\widehat{R}, \widehat{X}\right) \\ &= -\rho_{\hat{R},\hat{X}}\frac{\sqrt{V\left(\widehat{R}\right)}\sqrt{V\left(\widehat{X}\right)}}{X} \end{aligned} \tag{8.2.2}$$

where $\rho_{\hat{R},\hat{X}}$ is the correlation coefficient between $\widehat{R}$ and $\widehat{X}$.

Hence, the bias $B\left(\widehat{R}\right)$ is zero if the correlation between $\widehat{R}$ and $\widehat{X}$ is zero. In particular, if $y$ is approximately proportional to $x$, then $\widehat{R}$ is approximately constant and the bias $B\left(\widehat{R}\right)$ is negligible. Furthermore, since $\left|\rho_{\hat{R},\hat{X}}\right| \leq 1$, we have the following upper bound of the absolute relative bias of $\widehat{R}$ derived by Hartley and Ross (1954).

$$ARB\left(\widehat{R}\right) = \left|\frac{B\left(\widehat{R}\right)}{R}\right| \leq \frac{\sqrt{V\left(\widehat{R}\right)}}{|R|}\,\frac{\sqrt{V\left(\widehat{X}\right)}}{|X|} \tag{8.2.3}$$

$$\text{i.e., } \frac{\left|B\left(\widehat{R}\right)\right|}{\sqrt{V\left(\widehat{R}\right)}} \leq \left|C_{\hat{X}}\right| \tag{8.2.4}$$

where $C_{\hat{X}}$ is the coefficient of variation (CV) of $\widehat{X}$.

Eq. (8.2.4) indicates that if $C_{\hat{X}}$, is less than 0.1, the effect of bias is not an appreciable disadvantage in estimating $R$ (Cochran, 1977).

Let a sample $s$ of size $n$ be selected from a population $U$ of size $N$ with probability $p(s)$ and $\widehat{X}_s = \sum_{i\in s} b_{si}x_i$ and $\widehat{Y}_s = \sum_{i\in s} b_{si}y_i$ be the linear unbiased estimators for $X$ and $Y$, respectively. The constants $b_{si}$'s are free from $y_i$'s and satisfy the unbiasedness condition $\sum_{s\supset i} b_{si}p(s) = 1\ \forall i = 1, \ldots, N$. Consider the following estimator of $R$ based on an arbitrary sampling design $p$

$$\widehat{R} = \frac{\sum_{i\in s} b_{si}y_i}{\sum_{i\in s} b_{si}x_i} \tag{8.2.5}$$

Then, following Rao T.J. (1967a,b), we get the bias of $\widehat{R}$ as

$$B(\widehat{R}) = \sum_{s}\left(\frac{\sum_{i\in s} b_{si}y_i}{\sum_{i\in s} b_{si}x_i}\right)p(s) - R$$

$$= \frac{1}{X}\left[\sum_{s}\left(\sum_{i\in s} w_{si}y_i\right)p(s) - Y\right]$$

$$\left[\text{where } w_{si} = b_{si}X\Big/\left(\sum_{i\in s} b_{si}x_i\right)\right] \tag{8.2.6}$$

$$= \frac{1}{X}\left[\sum_{i\in U} y_i\left\{\sum_{s\supset i} w_{si}p(s) - 1\right\}\right]$$

$$= \frac{1}{X}\sum_{i\in U} y_i(\alpha_i - 1)$$

where $\alpha_i = \sum_{s\supset i} w_{si}p(s)$.

The mean-square error (MSE) of $\widehat{R}$ is given by

$$M(\widehat{R}) = E(\widehat{R} - R)^2 = \sum_{s}\left[\left(\frac{\sum_{i\in s} b_{si}y_i}{\sum_{i\in s} b_{si}x_i}\right)^2 - 2R\left(\frac{\sum_{i\in s} b_{si}y_i}{\sum_{i\in s} b_{si}x_i}\right) + R^2\right]$$

$$= \frac{1}{X^2}\left[\sum_{s}\left\{\left(\sum_{i\in s} w_{si}y_i\right)^2 - 2Y\left(\sum_{i\in s} w_{si}y_i\right)\right\}p(s) + Y^2\right]$$

$$= \frac{1}{X^2}\left[\sum_{i\in U}\beta_i y_i^2 + \sum_{i\neq}\sum_{j\in U}\beta_{ij}y_iy_j - 2\left(\sum_{i\in U}\alpha_i y_i\right)Y + Y^2\right]$$

$$\left(\text{where } \beta_i = \sum_{s\supset i} w_{si}^2 p(s) \text{ and } \beta_{ij} = \sum_{s\supset i,j} w_{si}w_{sj}p(s)\right)$$

$$= \frac{1}{X^2}\left[\sum_{i\in U}\gamma_i y_i^2 + \sum_{i\neq}\sum_{j\in U}\gamma_{ij}y_iy_j\right] \tag{8.2.7}$$

where $\gamma_i = \beta_i - 2\alpha_i + 1$, $\gamma_{ij} = \beta_{ij} - (\alpha_i + \alpha_j) + 1$.

From Eqs. (8.2.6) and (8.2.7), we set unbiased estimators for $B(\widehat{R})$ and $M(\widehat{R})$, when $X$ is known, as follows:

$$\widehat{B}(R)=\frac{1}{X}\sum_{i\in s}\frac{(\alpha_i-1)}{\pi_i}y_i \text{ and } \widehat{M}\left(\widehat{R}\right)=\frac{1}{X^2}\left(\sum_{i\in s}\frac{\gamma_i}{\pi_i}y_i^2+\sum_{i\neq}\sum_{j\in s}\frac{\gamma_{ij}}{\pi_{ij}}y_iy_j\right)$$

where $\pi_i$ and $\pi_{ij}(>0)$ are the inclusion probabilities of the $i$th unit and $i$th and $j$th unit, $i\neq j$.

In case $X$ is not known, we may replace $X$ by its unbiased estimate $\widehat{X}$. The results above are summarized in the following theorem:

### Theorem 8.2.1

Let $\widehat{R}=\dfrac{\sum_{i\in s}b_{si}y_i}{\sum_{i\in s}b_{si}x_i}$ be an estimator of $R$ based on a sampling design with inclusion probabilities $\pi_i$ and $\pi_{ij}(>0)$ for the $i$th and $i$th and $j$th with $b_{si}$'s are suitably chosen weights. Then,

(i) $B\left(\widehat{R}\right)=\dfrac{1}{X}\displaystyle\sum_{i\in U}y_i(\alpha_i-1)$

(ii) $M\left(\widehat{R}\right)=\left(\displaystyle\sum_{i\in U}\gamma_iy_i^2+\sum_{i\neq}\sum_{j\in U}\gamma_{ij}y_iy_j\right)\Big/X^2$

(iii) $\widehat{B}(R)=\dfrac{1}{X}\displaystyle\sum_{i\in s}\frac{(\alpha_i-1)}{\pi_i}y_i$

and

(iv) $\widehat{M}\left(\widehat{R}\right)=\dfrac{1}{X^2}\left(\displaystyle\sum_{i\in s}\frac{\gamma_i}{\pi_i}y_i^2+\sum_{i\neq}\sum_{j\in s}\frac{\gamma_{ij}}{\pi_{ij}}y_iy_j\right)$

where $\alpha_i=\sum_{s\supset i}w_{si}p(s), \beta_i=\sum_{s\supset i}w_{si}^2p(s), \beta_{ij}=\sum_{s\supset i,j}w_{si}w_{sj}p(s), \gamma_i=\beta_i-2\alpha_i+1, \gamma_{ij}=\beta_{ij}-(\alpha_i+\alpha_j)+1$ and $p(s)$ is the probability of selection of $s$.

### Remark 8.2.1

Although the expressions of bias, MSE, and their estimators, look elegant in Theorem 8.2.1, they cannot be used in practice because the expressions $\alpha_i$, $\beta_i$, $\beta_{ij}$ are difficult to compute even in simple random sampling without replacement (SRSWOR) sampling as pointed out by Sukhatme et al. (1984). For example, consider SRSWOR, where $p(s)=1\Big/\binom{N}{n}$, $b_{si}=N/n$, $w_{si}=b_{si}X\Big/\left(\sum_{i\in s}b_{si}x_i\right)=X/X_s$ and $X_s=\sum_{i\in s}x_i$.

Here we get

$$B\left(\widehat{R}\right)=\sum_{i\in U}y_i(\alpha_i\cdot-1),\ \widehat{B}\left(\widehat{R}\right)=\frac{1}{X}\sum_{i\in s}y_i(\alpha_i\cdot-1)/\pi_i,$$

$$M\left(\widehat{R}\right)=\frac{1}{X^2}\left(\sum_{i\in U}\gamma_i\cdot y_i^2+\sum_{i\neq}\sum_{j\in U}\gamma_{ij}\cdot y_iy_j\right)\text{and}$$

$$\widehat{M}\left(\widehat{R}\right)=\frac{1}{X^2}\left(\sum_{i\in s}\gamma_i\cdot y_i^2/\pi_i+\sum_{i\neq}\sum_{j\in s}\gamma_{ij}\cdot y_iy_j/\pi_{ij}\right)$$

where

$$\alpha_i\cdot=\sum_{s\supset i}\frac{X}{X_s}\Big/\binom{N}{n},\ \gamma_i\cdot=X^2\sum_{s\supset i}\frac{1}{X_s^2}-2X\sum_{s\supset i}\frac{1}{X_s}+1\text{ and}$$

$$\gamma_{ij}\cdot=X^2\sum_{s\supset i,j}\frac{1}{X_s^2}-X\left(\sum_{s\supset i}\frac{1}{X_s}+\sum_{s\supset j}\frac{1}{X_s}\right)+1$$

### 8.2.2 Approximate Expression of Bias and Mean-Square Errors

Following Murthy (1977), writing

$$\delta_y=\left(\widehat{Y}-Y\right)/Y\text{ and }\delta_x=\left(\widehat{X}-X\right)/X \tag{8.2.8}$$

we get $\widehat{R}=R(1+\delta_y)(1+\delta_x)^{-1}$.

Now assuming $|\delta_x|<1\ \forall\ s$ with $p(s)>0$, we can write $\widehat{R}$ as

$$\begin{aligned}\widehat{R}&=R(1+\delta_y)\left\{1-\delta_x+(\delta_x)^2-\cdots\right\}\\&=R\left[1+\{\delta_y-\delta_x\}+\left\{(\delta_x)^2-\delta_y\delta_x\right\}+\cdots\right]\end{aligned}\tag{8.2.9}$$

$$\text{i.e. }\widehat{R}-R=R\left[\{\delta_y-\delta_x\}+\left\{(\delta_x)^2-\delta_y\delta_x\right\}+\cdots\right]$$

For a large sample size $n$, $\eta_{ij}=E\left[(\delta_y)^i(\delta_x)^j\right]$ is expected to be small for $i+j>2$ (vide Murthy, 1977). Hence, neglecting $\eta_{ij}$ for $i+j>2$, we get the following expressions of the bias of $\widehat{R}$ up to the first order of approximation as follows:

$$\begin{aligned}
\text{Bias of } \widehat{R} = B(\widehat{R}) &\cong R\left[E\{\delta_y - \delta_x\} + E\{(\delta_x)^2 - \delta_y\delta_x\}\right] \\
&= R\left[V(\widehat{X})/X^2 - Cov(\widehat{X}, \widehat{Y})/(XY)\right] \\
&\quad (\text{since } E(\delta_x) = E(\delta_y) = 0) \\
&= \left[RV(\widehat{X}) - Cov(\widehat{X}, \widehat{Y})\right]/X^2 \\
&= R\left[C_{\hat{X}}^2 - \rho_{\hat{X},\hat{Y}} C_{\hat{X}} C_{\hat{Y}}\right]
\end{aligned} \tag{8.2.10}$$

where $C_{\hat{X}} = \sqrt{V(\widehat{X})/X}$ and $C_{\hat{Y}} = \sqrt{V(\widehat{Y})/Y}$ are the CVs of $\widehat{X}$ and $\widehat{Y}$, respectively and $\rho_{\hat{X},\hat{Y}}$ is the correlation coefficient between $\widehat{X}$ and $\widehat{Y}$.

$$\text{Hence,} \quad B(\widehat{R}) = -\left[Cov(\widehat{X}, \widehat{D})\right]/X^2 \tag{8.2.11}$$

(where $\widehat{D} = \widehat{Y} - R\widehat{X} = \sum_{i\in s} b_{si}d_i$ and $d_i = y_i - R\,x_i$)

$$= -\left[\sum_{i\in U} x_i d_i \theta_i + \sum_{i\neq}\sum_{j\in U} x_i d_j \theta_{ij}\right]\Big/ X^2 \tag{8.2.12}$$

where $\theta_i = \sum_{s\supset i} b_{si}^2 p(s)$ and $\theta_{ij} = \sum_{s\supset i,j} b_{si}b_{sj}p(s)$.

### Remark 8.2.2

If $d_i = y_i - R\ \ x_i = 0$ for every $i = 1,\ldots, N$, the regression of $y$ on $x$ passes through the origin and the bias $B(\widehat{R})$ is exactly equal to zero.

An approximate unbiased estimator for bias $B(\widehat{R})$ is given by

$$\begin{aligned}
\widehat{B}(\widehat{R}) &\cong \left[\widehat{R}\,\widehat{V}(\widehat{X}) - \widehat{C}ov(\widehat{X}, \widehat{Y})\right]\Big/\widehat{X}^2 \\
&= -\left[\sum_{i\in s} x_i \widehat{d}_i \theta_i \big/ \pi_i + \sum_{i\neq}\sum_{j\in s} x_i \widehat{d}_j \theta_{ij} \big/ \pi_{ij}\right]\Big/ \widehat{X}^2
\end{aligned} \tag{8.2.13}$$

where $\widehat{V}(\widehat{X})$ = unbiased estimator of $V(\widehat{X})$, $\widehat{C}ov(\widehat{X}, \widehat{Y})$ = unbiased estimator of $Cov(\widehat{X}, \widehat{Y})$ and $\widehat{d}_i = y_i - \widehat{R}\ \ x_i$.

From the expression (8.2.9), the expression of MSE of $\widehat{R}$ is obtained as

$$M(\widehat{R}) = E(\widehat{R} - R)^2 = R^2 E\left[(\delta_y - \delta_x) + \{(\delta_x)^2 - \delta_y\delta_x\} + \ldots\right]^2$$

If we neglect the terms $\eta_{ij} = E\left(\delta_x^i \delta_y^j\right)$ for $i + j > 2$, an approximate expression of $M\left(\widehat{R}\right)$ up to the first order of approximation is obtained as

$$M\left(\widehat{R}\right) \cong R^2\left[\frac{V\left(\widehat{Y}\right)}{Y^2} - 2\frac{Cov\left(\widehat{Y}, \widehat{X}\right)}{YX} + \frac{V\left(\widehat{X}\right)}{X^2}\right] \tag{8.2.14}$$

$$= R^2\left[C_{\hat{Y}}^2 - 2\rho_{\hat{X},\hat{Y}} C_{\hat{X}} C_{\hat{Y}} + C_{\hat{X}}^2\right]$$

$$= \left[V\left(\widehat{Y}\right) - 2R\,Cov\left(\widehat{Y}, \widehat{X}\right) + R^2\ V\left(\widehat{X}\right)\right]/X^2 \tag{8.2.15}$$

$$= V\left(\widehat{D}\right)/X^2 \tag{8.2.16}$$

$$= \left[\sum_{i\in U} d_i^2\theta_i + \sum_{i\neq}\sum_{j\in U} d_i d_j \theta_{ij}\right]\Big/ X^2 \tag{8.2.17}$$

(noting $D = \sum_{i\in U} d_i = 0$).

An approximate unbiased estimator for the mean square $M\left(\widehat{R}\right)$ is

$$\widehat{M}\left(\widehat{R}\right) \cong \left[\widehat{V}\left(\widehat{Y}\right) - 2\widehat{R}\,\widehat{C}ov\left(\widehat{Y}, \widehat{X}\right) + \widehat{R}^2\,\widehat{V}\left(\widehat{X}\right)\right]\Big/\widehat{X}^2 \tag{8.2.18}$$

(where $\widehat{V}\left(\widehat{Y}\right)$ is an unbiased estimator of $V\left(\widehat{Y}\right)$)

$$= \left[\sum_{i\in s} \widehat{d}_i^2\theta_i\Big/\pi_i + \sum_{i\neq}\sum_{j\in s} \widehat{d}_i \widehat{d}_j \theta_{ij}\Big/\pi_{ij}\right]\Big/ \widehat{X}^2 \tag{8.2.19}$$

All the estimators of biases and MSEs presented here are consistent for large samples. The results derived earlier have been summarized in the following theorem.

### Theorem 8.2.2

Approximate expressions of bias and MSE of $\widehat{R}$ up to the first order of approximation and their approximate unbiased estimates are as follows:

(i) $B\left(\widehat{R}\right) \cong \left[RV\left(\widehat{X}\right) - Cov\left(\widehat{X}, \widehat{Y}\right)\right]/X^2$

$$= -\left[Cov\left(\widehat{X}, \widehat{D}\right)\right]/X^2$$

$$= -\left(\sum_{i\in U} x_i d_i \theta_i + \sum_{i\neq}\sum_{j\in U} x_i d_j \theta_{ij}\right)\Big/ X^2$$

(ii) $M\left(\widehat{R}\right) \cong \left[V\left(\left(\widehat{Y}\right) - 2RCov\left(\widehat{X}, \widehat{Y}\right) + R^2 V\left(\widehat{X}\right)\right]\right/X^2$

$$= V\left(\widehat{D}\right)/X^2$$

$$= \left(\sum_{i \in U} d_i^2 \theta_i + \sum_{i \neq} \sum_{j \in U} d_i d_j \theta_{ij}\right) \Big/ X^2$$

$$= R^2\left(C_{\hat{Y}}^2 - 2\rho_{\hat{X},\hat{Y}} C_{\hat{X}} C_{\hat{Y}} + C_{\hat{X}}^2\right)$$

(iii) $\widehat{B}\left(\widehat{R}\right) = \left[\widehat{R}\,\widehat{V}\left(\widehat{X}\right) - \widehat{C}ov\left(\widehat{X}, \widehat{Y}\right)\right] \Big/ \widehat{X}^2$

$$= -\left(\sum_{i \in s} x_i \widehat{d}_i \theta_i \Big/ \pi_i + \sum_{i \neq} \sum_{j \in s} x_i \widehat{d}_j \theta_{ij} \Big/ \pi_{ij}\right) \Big/ \widehat{X}^2$$

and

(iv) $\widehat{M}\left(\widehat{R}\right) = \left[\widehat{V}\left(\widehat{Y}\right) - 2\widehat{R}\,\widehat{C}ov\left(\widehat{Y}, \widehat{X}\right) + \widehat{R}^2\,\widehat{V}\left(\widehat{X}\right)\right] \Big/ \widehat{X}^2$

$$= \left[\sum_{i \in s} \widehat{d}_i^2 \theta_i \Big/ \pi_i + \sum_{i \neq} \sum_{j \in s} \widehat{d}_i \widehat{d}_j \theta_{ij} \Big/ \pi_{ij}\right] \Big/ \widehat{X}^2$$

where $\widehat{V}\left(\widehat{Y}\right)$, $\widehat{V}\left(\widehat{X}\right)$, and $\widehat{C}ov\left(\widehat{Y}, \widehat{X}\right)$ are unbiased estimators of $V\left(\widehat{Y}\right)$, $V\left(\widehat{X}\right)$, and $Cov\left(\widehat{Y}, \widehat{X}\right)$; $\theta_i = \sum_{s \supset i} b_{si}^2 p(s)$, $\theta_{ij} = \sum_{s \supset i,j} b_{si} b_{sj} p(s)$, $d_i = y_i - R\ x_i$, and $\widehat{d}_i = y_i - \widehat{R}\ x_i$.

## 8.3 RATIO ESTIMATOR FOR POPULATION TOTAL

The ratio estimator for the population total $Y$ is defined as

$$\widehat{Y}_R = \frac{\widehat{Y}}{\widehat{X}} X = \widehat{R} X \tag{8.3.1}$$

where $X$ is the total of the auxiliary variable, which is known.

Clearly, the ratio estimator $\widehat{Y}_R$ is a biased estimator for the total $Y$ in general. The bias and MSE of $\widehat{Y}_R$ are obtained from the relations $B\left(\widehat{Y}_R\right) = XB\left(\widehat{R}\right)$ and $M\left(\widehat{Y}_R\right) = X^2 M\left(\widehat{R}\right)$, respectively. Approximate expressions of the bias and the MSE of the ratio estimator are directly obtained from Theorem 8.2.2 and have been given in the following theorem.

Theorem 8.3.1

The expressions of bias and MSE of $\widehat{Y}_R$ up to the first order of approximation and their approximate unbiased estimators are given by:

(i) $$B\left(\widehat{Y}_R\right) \cong \left[RV\left(\widehat{X}\right) - Cov\left(\widehat{X}, \widehat{Y}\right)\right]/X$$
$$= -\left[Cov\left(\widehat{X}, \widehat{D}\right)\right]/X$$
$$= -\left(\sum_{i\in U} x_i d_i \theta_i + \sum_{i\neq}\sum_{j\in U} x_i d_j \theta_{ij}\right) \Big/ X$$

(ii) $$M\left(\widehat{Y}_R\right) \cong \left[V\left(\widehat{Y}\right) - 2RCov\left(\widehat{X}, \widehat{Y}\right) + R^2 V\left(\widehat{X}\right)\right]$$
$$= V\left(\widehat{D}\right)$$
$$= -\left(\sum_{i\in U} d_i^2 \theta_i + \sum_{i\neq}\sum_{j\in U} d_i d_j \theta_{ij}\right)$$
$$= Y^2\left(C_{\hat{Y}}^2 - 2\rho_{\hat{X},\hat{Y}} C_{\hat{X}} C_{\hat{Y}} + C_{\hat{X}}^2\right)$$

(iii) $$\widehat{B}\left(\widehat{Y}_R\right) = \left[\widehat{R}\,\widehat{V}\left(\widehat{X}\right) - C\widehat{o}v\left(\widehat{X}, \widehat{Y}\right)\right]/\widehat{X}$$
$$= -\left(\sum_{i\in s} x_i \widehat{d}_i \theta_i \Big/ \pi_i + \sum_{i\neq}\sum_{j\in s} x_i \widehat{d}_j \theta_{ij} \Big/ \pi_{ij}\right) \Big/ \widehat{X}$$

and

(iv) $$\widehat{M}\left(\widehat{Y}_R\right) = \left[\widehat{V}\left(\widehat{Y}\right) - 2\widehat{R}\,C\widehat{o}v\left(\widehat{Y}, \widehat{X}\right) + \widehat{R}^2\,\widehat{V}\left(\widehat{X}\right)\right]$$
$$= \left(\sum_{i\in s} \widehat{d}_i^2 \theta_i \Big/ \pi_i + \sum_{i\neq}\sum_{j\in s} \widehat{d}_i \widehat{d}_j \theta_{ij} \Big/ \pi_{ij}\right)$$

## 8.3.1 Efficiency of the Ratio Estimator

The MSE of $\widehat{Y}_R$ becomes zero if $d_i = y_i - R\,x_i = 0$ for every $i = 1,\ldots, N$. Hence the ratio estimator becomes optimal if the study variable $y$ is exactly proportional to the auxiliary variable. However, in practice, it is not possible to have $d_i$ exactly equal to zero for every $i = 1,\ldots, N$. But it is quite possible that study variable $y$ is approximately proportional to the auxiliary variable $x$. In such cases, the use of auxiliary information $x$ makes the ratio estimator highly efficient by reducing the magnitude of the mean square substantially.

The ratio estimator $\widehat{Y}_R$ becomes more efficient than the conventional estimator $\widehat{Y}$ if

$$
\begin{aligned}
&V\left(\widehat{Y}\right) - M\left(\widehat{Y}_R\right) \geq 0 \\
&\text{i.e., } 2RCov\left(\widehat{Y}, \widehat{X}\right) - R^2 V\left(\widehat{X}\right) \geq 0
\end{aligned} \tag{8.3.2}
$$

$$
\begin{aligned}
\text{i.e., } \rho_{\hat{X}\hat{Y}} &> \frac{1}{2}\left(\sqrt{V\left(\widehat{X}\right)}\Big/X\right)\Big/\left(\sqrt{V\left(\widehat{Y}\right)}\Big/Y\right) \\
&= \frac{1}{2}\left(C_{\hat{X}}/C_{\hat{Y}}\right) \quad \text{when } R > 0
\end{aligned} \tag{8.3.3}
$$

$$
\begin{aligned}
\text{or } \rho_{\hat{X}\hat{Y}} &< -\frac{1}{2}\left(\sqrt{V\left(\widehat{X}\right)}\Big/X\right)\Big/\left(\sqrt{V\left(\widehat{Y}\right)}\Big/Y\right) \\
&-\frac{1}{2}\left|C_{\hat{X}}/C_{\hat{Y}}\right| \quad \text{when } R < 0
\end{aligned} \tag{8.3.4}
$$

### 8.3.2 Optimality of the Ratio Estimator

It is pointed out in Theorem 6.4.2 that the ratio estimator $T_{01} = \widehat{Y}_R$ is the optimal in the class of linear $\xi$-unbiased predictors $C_{l\xi}$ of the population $Y$, under the superpopulation model $E_\xi(y_i) = \beta x_i$, $V_\xi(y_i) = \sigma^2 x_i$, and $C_\xi(y_i, y_j) = 0$ for $i \neq j$, i.e., the regression of $y$ on $x$ passes through the origin and the variance of $y_i$ is proportional to $x_i$.

## 8.4 BIASES AND MEAN-SQUARE ERRORS FOR SPECIFIC SAMPLING DESIGNS

In this section, the expressions of biases, MSEs, and their approximate unbiased estimators for a few well-known sampling designs, which are commonly used in practice, are given. The conditions of superiority of the ratio estimator over the conventional estimators are also given. These results are derived as special cases of the results presented in Sections 8.2 and 8.3.

### 8.4.1 Fixed Effective Sample Size ($n$) Design

Let $b_{si} = y_i/\pi_i$, then we get

$$
\widehat{Y} = \sum_{i\in s} y_i/\pi_i,\ \widehat{X} = \sum_{i\in s} x_i/\pi_i,\ \widehat{R} = \frac{\sum_{i\in s} y_i/\pi_i}{\sum_{i\in s} x_i/\pi_i},\ \widehat{Y}_R = \frac{\sum_{i\in s} y_i/\pi_i}{\sum_{i\in s} x_i/\pi_i} X
$$

Now noting $\theta_i = \sum_{s \supset i} b_{si}^2 p(s) = 1/\pi_i$ and $\theta_{ij} = \sum_{s \supset i,j} b_{si} b_{sj} p(s) = \pi_{ij}/(\pi_i \pi_j)$, the approximate expressions of the bias, MSE, and their unbiased estimators are obtained as follows:

$$B(\widehat{Y}_R) = -\left[\sum_i x_i d_i/\pi_i + \sum_{i \neq} \sum_j x_i d_j \pi_{ij}/\pi_i \pi_j\right] \Big/ X \quad (8.4.1)$$

$$\begin{aligned} M(\widehat{Y}_R) &= \left(\sum_{i \in U} \frac{d_i^2}{\pi_i} + \sum_{i \neq} \sum_{j \in U} d_i d_j \frac{\pi_{ij}}{\pi_i \pi_j}\right) \\ &= \frac{1}{2} \sum_{i \neq} \sum_{j \in U} (\pi_i \pi_j - \pi_{ij}) \left(\frac{d_i}{\pi_i} - \frac{d_j}{\pi_j}\right)^2 \end{aligned} \quad (8.4.2)$$

$$\widehat{B}(\widehat{Y}_R) = -\left[\sum_{i \in s} x_i \widehat{d}_i \Big/ \pi_i^2 + \sum_{i \neq} \sum_{j \in s} x_i \widehat{d}_j \Big/ (\pi_i \cdot \pi_j)\right] \Big/ \widehat{X} \quad (8.4.3)$$

and

$$\widehat{M}(\widehat{Y}_R) = \frac{1}{2} \sum_{i \neq} \sum_{j \in s} \frac{(\pi_i \pi_j - \pi_{ij})}{\pi_{ij}} \cdot \left(\frac{\widehat{d}_i}{\pi_i} - \frac{\widehat{d}_j}{\pi_j}\right)^2 \quad (8.4.4)$$

### 8.4.2 Simple Random Sampling Without Replacement

For SRSWOR $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/\{N(N-1)\}$; hence, for the choice of $b_{si} = 1/\pi_i = N/n$, we get $\widehat{Y} = N\overline{y}_s$, $\widehat{X} = N\overline{x}_s$, $\widehat{R} = \frac{\overline{y}_s}{\overline{x}_s}$, $\widehat{Y}_R = \frac{\overline{y}_s}{\overline{x}_s} X$, where $\overline{y}_s = \sum_{i \in s} y_i/n$ and $\overline{x}_s = \sum_{i \in s} x_i/n$.

Theorem 8.3.1 yields

$$\begin{aligned} B(\widehat{Y}_R) &= \left[RV(\widehat{X}) - Cov(\widehat{X}, \widehat{Y})\right]/X \\ &= N^2 \left(\frac{1}{n} - \frac{1}{N}\right) \left(RS_x^2 - \rho S_x S_y\right)/X \end{aligned} \quad (8.4.5)$$

$$M(\widehat{Y}_R) = N^2 \left(\frac{1}{n} - \frac{1}{N}\right) \left(S_y^2 - 2R\rho S_x S_y + R^2 S_x^2\right) \quad (8.4.6)$$

$$= N^2 \left(\frac{1}{n} - \frac{1}{N}\right) S_d^2 \quad (8.4.7)$$

where $(N-1)S_y^2 = \sum_{i \in U} (y_i - \overline{Y})^2$, $(N-1)S_x^2 = \sum_{i \in U} (x_i - \overline{X})^2$,

$(N-1)S_d^2 = \sum_{i\in U} d_i^2$, $\overline{X} = X/N$, $\overline{Y} = Y/N$, $\rho = S_{xy}/(S_x S_y) =$ population correlation coefficient between $x$ and $y$, and
$(N-1)S_{xy} = \sum_{i\in U}\left(y_i - \overline{Y}\right)\left(x_i - \overline{X}\right)$.

Approximate unbiased estimators for $B\left(\widehat{Y}_R\right)$ and $M\left(\widehat{Y}_R\right)$ are also obtained from Theorem 8.3.1 as follows:

$$\begin{aligned}\widehat{B}\left(\widehat{Y}_R\right) &= N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left(\widehat{R}s_x^2 - s_{xy}\right)/\widehat{X}\\ \widehat{M}\left(\widehat{Y}_R\right) &= N^2\left(\frac{1}{n}-\frac{1}{N}\right)\left(s_y^2 - 2\widehat{R}s_{xy} + \widehat{R}^2 s_x^2\right)\\ &= N^2\left(\frac{1}{n}-\frac{1}{N}\right)s_{\hat{d}}^2\end{aligned} \tag{8.4.8}$$

where $(n-1)s_y^2 = \sum_{i\in s}\left(y_i - \overline{y}_s\right)^2$, $(n-1)s_x^2 = \sum_{i\in s}\left(x_i - \overline{x}_s\right)^2$, $(n-1)s_{\hat{d}}^2 = \sum_{i\in s}\widehat{d}_i^2$, and $(n-1)s_{xy} = \sum_{i\in s}\left(y_i - \overline{y}_s\right)\left(x_i - \overline{x}_s\right)$.

Finally, using Eqs. (8.3.3) and (8.3.4), we get conditions of superiority of the ratio estimator $\widehat{Y}_R = \dfrac{\overline{y}_s}{\overline{x}_s}X$ over the conventional estimator $N\overline{y}_s$ for the total $Y$ as

$$\rho > \frac{1}{2}\left(C_x/C_y\right) \text{ with } R > 0 \text{ and } \rho < -\frac{1}{2}\left|C_x/C_y\right| \text{ with } R < 0 \tag{8.4.9}$$

where $C_x$ and $C_y$ denote the CVs of $x$ and $y$, respectively. In the case where the CVs of $x$ and $y$ are equal, i.e., $C_x = C_y$, the ratio estimator fares better if $\rho > 0.5$ with $R > 0$ and $\rho < -0.5$ with $R < 0$.

### 8.4.3 Probability Proportional to Size With Replacement

Let the sample $s$ be selected by PPSWR sampling scheme using $z_i(>0)$ as the known size measure for the $i$th unit and $p_i = z_i/Z$ $\left(Z = \sum_{i\in U} z_i\right)$ as its normed size measure. In this case,

$$\widehat{Y} = \sum_i \frac{n_i(s)y_i}{np_i},\ \widehat{X} = \sum_i \frac{n_i(s)x_i}{np_i},\ \widehat{R} = \frac{\sum_i n_i(s)y_i/p_i}{\sum_i n_i(s)x_i/p_i};$$

$$\widehat{Y}_R = \left(\frac{\sum_i n_i(s)y_i/p_i}{\sum_i n_i(s)x_i/p_i}\right)X$$

where $n_i(s)$ denotes the number of times $i$th unit appears in the sample $s$.

Now writing $\sum_{i \in U} p_i \left(\frac{y_i}{p_i} - Y\right)^2 = V_{pyy}$, $\sum_{i \in U} p_i \left(\frac{x_i}{p_i} - X\right)^2 = V_{pxx}$, $\sum_{i \in U} p_i \left(\frac{x_i}{p_i} - X\right)\left(\frac{y_i}{p_i} - Y\right) = V_{pxy}$, and $\sum_{i \in U} \frac{d_i^2}{p_i} = V_{pdd}$, we get

$$B\left(\widehat{Y}_R\right) = \left[RV\left(\widehat{X}\right) - Cov\left(\widehat{X}, \widehat{Y}\right)\right]/X = (RV_{pxx} - V_{pxy})/(nX) \quad (8.4.10)$$

$$M\left(\widehat{Y}_R\right) = (V_{pyy} - 2RV_{pxy} + R^2 V_{pxx})/n \quad (8.4.11)$$

$$= V_{pdd}/n \quad (8.4.12)$$

Approximate unbiased estimators of $B\left(\widehat{Y}_R\right)$ and $M\left(\widehat{Y}_R\right)$ come out as follows:

$$\widehat{B}\left(\widehat{Y}_R\right) = \left(\widehat{R}\,\widehat{V}_{pxx} - \widehat{V}_{pxy}\right)/\left(n\widehat{X}\right) \quad (8.4.13)$$

$$\widehat{M}\left(\widehat{Y}_R\right) = \left(\widehat{V}_{pyy} - 2\widehat{R}\,\widehat{V}_{pxy} + \widehat{R}^2\,\widehat{V}_{pxx}\right)/n \quad (8.4.14)$$

$$= \widehat{V}_{p\hat{d}\hat{d}}\Big/n \quad (8.4.15)$$

where $\widehat{V}_{pxx} = \sum_i n_i(s)\left(\frac{x_i}{p_i} - \widehat{X}\right)^2\Big/(n-1)$,

$\widehat{V}_{pyy} = \sum_i n_i(s)\left(\frac{y_i}{p_i} - \widehat{Y}\right)^2\Big/(n-1)$,

$\widehat{V}_{pxy} = \sum_i n_i(s)\left(\frac{x_i}{p_i} - \widehat{X}\right)\left(\frac{y_i}{p_i} - \widehat{Y}\right)\Big/(n-1)$,

$\widehat{V}_{p\hat{d}\hat{d}} = \sum_i n_i(s)\left(\frac{\hat{d}_i}{p_i} - \widehat{D}\right)^2\Big/(n-1)$, $\widehat{D} = \sum_i \frac{n_i(s)\hat{d}_i}{np_i}$, and $\hat{d}_i = y_i - \widehat{R}x_i$.

### 8.4.4 Simple Random Sampling With Replacement

PPSWR reduces to simple random sampling with replacement, when $p_i = 1/N$ for every $i = 1,\ldots, N$. So, putting $p_i = 1/N$ in Section 8.4.3, we get the following:

$\widehat{Y} = N\overline{y}_n$, $\widehat{X} = N\overline{x}_n$, $\widehat{R} = \overline{y}_n/\overline{x}_n$, $\widehat{Y}_R = \widehat{Y}_R = \left(\overline{y}_n/\overline{x}_n\right)X$, $V\left(\widehat{Y}\right) = N(N-1)S_y^2\Big/n$, $V\left(\widehat{X}\right) = N(N-1)S_x^2/n$, and $Cov\left(\widehat{Y}, \widehat{X}\right) = N(N-1)S_{xy}/n$, where $\overline{x}_n$ and $\overline{y}_n$ denote the sample means of $x$ and $y$ based on all $n$ units including repetition.

$$B\left(\widehat{Y}_R\right) = \left[RV\left(\widehat{X}\right) - Cov\left(\widehat{X}, \widehat{Y}\right)\right]/X \\ = (N-1)\left(RS_x^2 - \rho S_x S_y\right)/\left(n\overline{X}\right) \tag{8.4.16}$$

$$M\left(\widehat{Y}_R\right) = N(N-1)\left(S_y^2 - 2R\rho S_x S_y + R^2 S_x^2\right)\Big/ n \tag{8.4.17}$$

$$= N(N-1)S_d^2/n \tag{8.4.18}$$

Approximate unbiased estimator of $B\left(\widehat{Y}_R\right)$ and $M\left(\widehat{Y}_R\right)$ are given as follows:

$$\widehat{B}\left(\widehat{Y}_R\right) = (N-1)\left[\widehat{R}s_x^2 - \widehat{\rho}s_x s_y\right]/(n\,\overline{x}_s) \tag{8.4.19}$$

$$\widehat{M}\left(\widehat{Y}_R\right) = N(N-1)\left[s_y^2 - 2\widehat{R}s_{xy} + \widehat{R}^2 s_x^2\right]\Big/ n \tag{8.4.20}$$

$$= N(N-1)s_{\hat{d}}^2\Big/ n \tag{8.4.21}$$

where $(n-1)s_x^2 = \sum_i n_i(s)\{x_i - \overline{x}(s)\}^2$, $(n-1)s_y^2 = \sum_i n_i(s)\left\{y_i - \overline{y}(s)\right\}^2$, $(n-1)s_{\hat{d}}^2 = \sum_i n_{is}\widehat{d}_i^2$, $(n-1)s_{xy} = \sum_i n_i(s)\{x_i - \overline{x}(s)\}\left\{y_i - \overline{y}(s)\right\}$, and $\widehat{\rho} = s_{xy}/(s_x\, s_y)$.

### Remark 8.4.1

The MSE of $\widehat{Y}_R$ can be estimated by using the standard Jackknife procedure. Details have been given in Chapter 18. The estimators $\widehat{M}_1\left(\widehat{Y}_R\right) = X^2\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{\overline{X}}{\overline{x}}\right)^2 s_{dy}^2$ and $\widehat{M}_2\left(\widehat{Y}_R\right) = X^2\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{\overline{X}}{\overline{x}}\right)s_{dy}^2$ were proposed by Rao (1969) and Wu (1982), respectively. Various authors including Rao and Rao (1971), Rao (1981), and Kreweski and Chakrabarti (1981) proposed alternative variance estimators and studied their performances using superpopulation models. None of the estimators was found to be the best in all the situations.

## 8.5 INTERVAL ESTIMATION

If the sample size $n(>35)$ is large, the distribution of $\widehat{Y}_R$ is asymptotically normal, and the confidence interval of $Y$ is obtained by applying the central limit theorem. Hence, $(1-\alpha)\times 100\%$ confidence interval for the total $Y$ is given by

$$\widehat{Y}_R \pm z_{\alpha/2}\sqrt{\widehat{M}\left(\widehat{Y}_R\right)} \tag{8.5.1}$$

where $z_{\alpha/2}$ is the upper $100\alpha/2$ point of a standard normal distribution.

For SRSWOR the confidence interval for $Y$ and $R$ are obtained, respectively, using the following formulae:

$$N\left[\overline{y}_s \pm z_{\alpha/2}\sqrt{\frac{(1-f)}{n}s_{\hat{d}}^2}\right] \tag{8.5.2}$$

and

$$\widehat{R} \pm z_{\alpha/2}\sqrt{\frac{(1-f)}{n}s_{\hat{d}}^2\Big/(\overline{x}_s)^2} \tag{8.5.3}$$

If the sample size $n$ is small, we cannot approximate the distribution of $\widehat{Y}_R$ as normal in general. In this situation, if we assume that $(x_1, y_1),\ldots, (x_n, y_n)$ are random samples from a bivariate normal distribution, then

$$t_{n-1} = \frac{\left(\overline{y}_s - R\overline{x}_s\right)}{\sqrt{(1-f)\left(s_y^2 - 2Rs_{xy} + R^2 s_x^2\right)\Big/n}}$$

follows $t$ distribution with $n-1$ degrees of freedom.

The $(1-\alpha)\times 100\%$ confidence interval of $R$ is $(R_1, R_2)$, where $R_1$ and $R_2$ are the roots of the quadratic equation

$$\left(\overline{y}_s - R\overline{x}_s\right)^2 - t_{\alpha/2,n-1}^2(1-f)\left(s_y^2 - 2Rs_{xy} + R^2 s_x^2\right)\Big/n = 0 \tag{8.5.4}$$

and $t_{\alpha/2,n-1}$ is the upper $100\alpha/2$ point of $t$ distribution with $n-1$ degrees of freedom.

### Example 8.5.1

A sample of 15 orchards was selected from 125 orchards of a certain locality by SRSWOR method.

Table 8.5.1 shows the number of plants and production of apples in the selected orchards.

(i) Estimate the average yield of apple per plant and its standard error.

(ii) Estimate the mean yield of apple per orchard and obtain 90% confidence interval of the mean yield when it is known that the average number of plants per orchard is 40 kg. Estimate the relative efficiency of the ratio estimator with respect to the sample mean.

**Table 8.5.1** Production of apples in selected orchards

| **Orchards** | **Number of plants** | **Production of apples (kg)** |
|---|---|---|
| 1 | 25 | 740 |
| 2 | 30 | 1000 |
| 3 | 50 | 1450 |
| 4 | 15 | 400 |
| 5 | 30 | 950 |
| 6 | 60 | 1700 |
| 7 | 30 | 1000 |
| 8 | 40 | 1260 |
| 9 | 45 | 1300 |
| 10 | 20 | 650 |
| 11 | 40 | 1250 |
| 12 | 30 | 850 |
| 13 | 50 | 1400 |
| 14 | 40 | 1250 |
| 15 | 20 | 550 |

Let $x_i$ and $y_i$, respectively, be the number of the apple plants ($x$) and production of apples ($y$) for the $i$th orchard, $i = 1, \ldots, 15$. Here, $n = 15$, $N = 125$, $\overline{X} = 40$, $\overline{x}(s) = 35$, and $\overline{y}(s) = 1050$.

Hence, estimated yield per plant $= \widehat{R} = \overline{y}(s)/\overline{x}(s) = 30$ kg. Estimated MSE of $\widehat{R} = \widehat{M}\left(\widehat{R}\right) = (1/n - 1/N)s_{\hat{d}}^2 \Big/ \overline{X}^2 = 0.173$, since $s_{\hat{d}}^2 = \dfrac{1}{n-1}\sum_{i\in s}\left(y_i - \widehat{R}x_i\right)^2 = 4728.571.$

Estimated standard error of $\widehat{R} = SE\left(\widehat{R}\right) = \sqrt{\widehat{M}\left(\widehat{R}\right)} = 0.416.$

The ratio estimate of the mean yield $\left(\overline{Y}\right)$ of apple per orchard is $\widehat{\overline{Y}}_R = 30\overline{X} = 1200$. Estimated standard error of $\widehat{\overline{Y}}_R = SE\left(\widehat{\overline{Y}}_R\right) = \overline{X}\sqrt{\widehat{M}\left(\widehat{R}\right)} = 16.655.$

Assuming that sample size is large, a 90% confidence interval for the mean yield is $\widehat{\overline{Y}}_R \pm t_{0.05,14}\overline{X}\sqrt{(1-f)s_{\hat{d}}^2 \big/ n} = 1200 \pm 1.761 \times 16.655 = (1170.670, 1229.331).$

The estimated efficiency of the ratio estimator $\widehat{\overline{Y}}_R = \dfrac{\hat{V}\{\overline{y}(s)\}}{\hat{M}\left(\hat{\overline{Y}}_R\right)}100 = \dfrac{s_y^2}{s_d^2}100 = 133942.9 \times 100/4728.571 = 2832.6\%.$

## 8.6 UNBIASED RATIO, ALMOST UNBIASED RATIO, AND UNBIASED RATIO–TYPE ESTIMATORS

### 8.6.1 Unbiased Ratio Estimator

It is shown in Section 5.5 that the estimator $\widehat{Y}_{lms} = \dfrac{\sum_{i\in s} y_i}{\sum_{i\in s} x_i} X$ is unbiased for the total $Y$ under the Lahiri-Midzuno-Sen (1951, 1952, 1953) sampling scheme. The estimator $\widehat{Y}_{lms}$ is known as the unbiased ratio estimator. The details properties of $\widehat{Y}_{lms}$ have been discussed in Section 5.5.

### 8.6.2 Almost Unbiased Ratio Estimator

Let $k$-independent samples, each of size $m = n/k$, be selected using the same sampling design. Let $\widehat{Y}_i$ and $\widehat{X}_i$ be the unbiased estimators of $Y$ and $X$, respectively, based on the $i$th sample $s_i$. Let the ratio estimator computed from the sample $s_i$ be

$$\widehat{Y}_R(i) = \frac{\widehat{Y}_i}{\widehat{X}_i} X$$

From the samples $s_i$'s, we can find two estimators of $Y$ as follows:

$$\widehat{Y}_{\overline{R}} = \frac{1}{k}\sum_{i=1}^{K} \widehat{Y}_R(i) \text{ and } \widehat{Y}_R(c) = \frac{\left(\sum_{i=1}^{k} \widehat{Y}_i/k\right)}{\left(\sum_{i=1}^{k} \widehat{X}_i/k\right)} X$$

Since all the samples $s_i$'s are of the same size $m$ and are selected by the same sampling design, we have approximate expressions of biases of $\widehat{Y}_R(i)$, $\widehat{Y}_{\overline{R}}$, and $\widehat{Y}_R(c)$ as follows:

$$B\left(\widehat{Y}_R(i)\right) = \left[RV\left(\widehat{X}_i\right) - Cov\left(\widehat{X}_i, \widehat{Y}_i\right)\right]/X = b(\text{say})\text{for } \forall\, i = 1, \ldots, k,$$
$$B\left(\widehat{Y}_{\overline{R}}\right) = b,$$

and

$$\begin{aligned} B\left(\widehat{Y}_R(c)\right) &= \left[RV\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{X}_i\right) - Cov\left(\frac{1}{k}\sum_{i=1}^{k}\widehat{X}_i, \frac{1}{k}\sum_{i=1}^{k}\widehat{Y}_i\right)\right]\Big/ X \\ &= \left[\sum_{i=1}^{k}\left\{RV\left(\widehat{X}_i\right) - Cov\left(\widehat{X}_i, \widehat{Y}_i\right)\right\}\right]\Big/ (k^2 X) \\ &= b/k \end{aligned} \tag{8.6.1}$$

Eq. (8.6.1) yields $E\left(\widehat{Y}_{\overline{R}} - \widehat{Y}_R(c)\right) = B\left(\widehat{Y}_{\overline{R}}\right) - B\left(\widehat{Y}_R(c)\right) = b\dfrac{(k-1)}{k}$. Hence an unbiased estimator of $b$ is

$$\widehat{b} = \frac{k}{k-1}\left(\widehat{Y}_{\overline{R}} - \widehat{Y}_R(c)\right) \tag{8.6.2}$$

Finally, subtracting an estimate of the amount of bias from $\widehat{Y}_R(c)$, we get an almost unbiased ratio estimator of $Y$ as

$$\begin{aligned}\widehat{Y}_{RAU} &= \widehat{Y}_R(c) - \widehat{B}\left(\widehat{Y}_R(c)\right)\\ &= \widehat{Y}_R(c) - \frac{1}{k-1}\left(\widehat{Y}_{\overline{R}} - \widehat{Y}_R(c)\right)\\ &= \frac{k\widehat{Y}_R(c) - \widehat{Y}_{\overline{R}}}{k-1}\end{aligned} \tag{8.6.3}$$

Under SRSWOR sampling, $\widehat{Y}_{\overline{R}} = \dfrac{1}{k}\displaystyle\sum_{i=1}^{k}\frac{\overline{y}_i}{\overline{x}_i}X$ and

$\widehat{Y}_R(c) = \dfrac{\sum_{i=1}^{k}\overline{y}_i}{\sum_{i=1}^{k}\overline{x}_i}X = \dfrac{\overline{y}}{\overline{x}}X$, where $\overline{y}_i$ and $\overline{x}_i$, respectively, denote sample means of $y$ and $x$ for the $i$th group, whereas $\overline{y}$ and $\overline{x}$, respectively, denote the sample means for $y$ and $x$ based on all the sampled units. It can be easily checked that under SRSWOR, the estimator (8.6.3) reduces to Murthy-Nanjamma (1957) estimator

$$\widehat{Y}_{MN} = \frac{k}{k-1}\frac{\overline{y}}{\overline{x}}X - \frac{1}{k-1}\left(\frac{1}{k}\sum_{i=1}^{k}\frac{\overline{y}_i}{\overline{x}_i}\right)X \tag{8.6.4}$$

### 8.6.3 Unbiased Ratio—Type Estimators

Let a sample $s$ of size $n$ be selected from a population with probability $p(s)$ and let $\widehat{Y} = \sum_{i\in s} b_{si}y_i$ be an unbiased estimator of $Y$. The constants $b_{si}$'s satisfy the unbiasedness condition $\sum_{s\supset i} b_{si}p(s) = 1$. Let $r_i = y_i/x_i$ and $\widehat{Y}_r = X\sum_{i\in s} b_{si}r_i/N$. Then,

$$E\left(\widehat{Y}_r\right) = \frac{X}{N}\sum_{i} r_i \sum_{s\supset i} b_{si}p(s) = \frac{X}{N}\sum_{i\in U} r_i \text{ as } \sum_{s\supset i} b_{si}p(s) = 1$$

and the bias of $\widehat{Y}_r$ is

$$B\left(\widehat{Y}_r\right) = E\left(\widehat{Y}_r\right) - Y = -\sum_{i\in U} r_i\left(x_i - \overline{X}\right)$$
$$= -\left[(N-1)\sum_{i\in U} r_i x_i - \sum_{i\neq}\sum_{j\in U} r_i x_j\right]\Big/ N \tag{8.6.5}$$

An exact unbiased estimator of $B\left(\widehat{Y}_r\right)$ is

$$\widehat{B}\left(\widehat{Y}_r\right) = -\left[(N-1)\sum_{i\in s} r_i x_i/\pi_i - \sum_{i\neq}\sum_{j\in s} r_i x_j/\pi_{ij}\right]\Big/ N$$

Hence the estimator

$$\widehat{Y}_r - \widehat{B}\left(\widehat{Y}_r\right) = \widehat{Y}_r + \left[(N-1)\sum_{i\in s} r_i x_i/\pi_i - \sum_{i\neq}\sum_{j\in s} r_i x_j/\pi_{ij}\right]\Big/ N$$
$$= \widehat{Y}_{RU} \tag{8.6.6}$$

is exactly unbiased for the total $Y$. The estimator $\widehat{Y}_{RU}$ is known as an unbiased ratio-type estimator.

### 8.6.4 Hartley—Ross Estimator

For SRSWOR sampling, $\pi_i = n/N, \pi_{ij} = n(n-1)/\{N(N-1)\}$. If we choose $b_{si} = 1/\pi_i = N/n$, then Eq. (8.6.6) reduces to

$$\widehat{Y}_{RU} = \widehat{Y}_{HR} = \overline{r}_s X + \frac{n(N-1)}{n-1}\left(\overline{y}_s - \overline{r}_s\overline{x}_s\right) \tag{8.6.7}$$

where $\overline{r}_s = \sum_{i\in s} r_i/n$.

The estimator $\widehat{Y}_{HR}$ is the well-known Hartley—Ross (1954) estimator. Several authors proposed alternative ratio-type unbiased and almost unbiased estimators, e.g., Goodman and Hartley (1958), Murthy and Nanjamma (1957), De Pascal (1961), Beale (1962), Tin (1965), Rao T.J. (1966, 1967a,b), among others. As far as efficiency (MSE) is concerned, none of the proposed alternative estimators are unconditionally superior to the conventional ratio estimator $\widehat{Y}_R$. The general method of bias reduction using Jackknife technique has been given in Chapter 18.

## 8.7 RATIO ESTIMATOR FOR STRATIFIED SAMPLING

Consider a population stratified into $K$ strata and that $X_i$ and $Y_i$ are, respectively, the total of the auxiliary variable $x$ and the study variable $y$ for the $i$th stratum $i = 1,\ldots, K$. From the $i$th stratum of size $N_i$, a sample $s_i$ of size $n_i$ is selected by some suitable sampling design. Let $\widehat{X}_i$ and $\widehat{Y}_i$ denote the estimators of $X_i$ and $Y_i$ based on the sample $s_i$. In case $X_i$ is known, the ratio estimator for the $i$th stratum total $Y_i$ is given by

$$\widehat{Y}_{iR} = \frac{\widehat{Y}_i}{\widehat{X}_i} X_i \tag{8.7.1}$$

Hence, ratio estimator for the population total $Y$ is given by

$$\widehat{Y}_{SR} = \sum_{i=1}^{K} \widehat{Y}_{iR} = \sum_{i=1}^{K} \frac{\widehat{Y}_i}{\widehat{X}_i} X_i \tag{8.7.2}$$

The ratio estimator (8.7.2) is termed as "separate ratio estimator."

Since $\widehat{Y}_{st} = \sum_{i=1}^{K} \widehat{Y}_i$ and $\widehat{X}_{st} = \sum_{i=1}^{K} \widehat{X}_i$ are unbiased for the population totals $Y$ and $X$, we define an alternative ratio estimator, which is called a "combined ratio estimator" for the total $Y$ as

$$\widehat{Y}_{CR} = \frac{\widehat{Y}_{st}}{\widehat{X}_{st}} X = \widehat{R}_C X \tag{8.7.3}$$

where $\widehat{R}_C = \dfrac{\widehat{Y}_{st}}{\widehat{X}_{st}}$.

### 8.7.1 Separate Ratio Estimator

Now noting the bias of $\widehat{Y}_{SR} = B\left(\widehat{Y}_{SR}\right) = \sum_{i=1}^{K} B\left(\widehat{Y}_{iR}\right)$ and the MSE of $\widehat{Y}_{SR} = M\left(\widehat{Y}_{SR}\right) = \sum_{i=1}^{K} M\left(\widehat{Y}_{iR}\right)$, we get the following result straight from Theorem 8.3.1 as follows:

#### Theorem 8.7.1

Approximate expressions of bias, MSE of $\widehat{Y}_{SR}$ (up to the first order of approximation) and their approximate unbiased estimators are given by:

(i) $B\left(\widehat{Y}_{SR}\right) \cong \sum_{i=1}^{K} \left(R_i V\left(\widehat{X}_i\right) - Cov\left(\widehat{X}_i, \widehat{Y}_i\right)\right) \Big/ X_i$

(ii) $M\left(\widehat{Y}_{SR}\right) \cong \sum_{i=1}^{K} \left(V\left(\widehat{Y}_i\right) - 2R_i Cov\left(\widehat{Y}_i, \widehat{X}_i\right) + R_i^2 V\left(\widehat{X}_i\right)\right)$

(iii) $\widehat{B}\left(\widehat{Y}_{SR}\right) \cong \sum_{i=1}^{K} \left(\widehat{R}_i \widehat{V}\left(\widehat{X}_i\right) - \widehat{C}ov\left(\widehat{X}_i, \widehat{Y}_i\right)\right) \Big/ \widehat{X}_i$

(iv) $\widehat{M}\left(\widehat{Y}_{SR}\right) \cong \sum_{i=1}^{K} \left(\widehat{V}\left(\widehat{Y}_i\right) - 2\widehat{R}_i Cov\left(\widehat{Y}_i, \widehat{X}_i\right) + R_i^2 V\left(\widehat{X}_i\right)\right)$

where $R_i = Y_i / X_i$ and $\widehat{Cov}\left(\widehat{X}_i, \widehat{Y}_i\right)$ is an unbiased estimator of $Cov\left(\widehat{X}_i, \widehat{Y}_i\right)$.

### 8.7.2 Combined Ratio Estimator

Theorem 8.3.1 yields approximate expressions of bias and MSE of $\widehat{Y}_{CR}$ as well as their approximate unbiased estimators as follows:

**Theorem 8.7.2**

$$B\left(\widehat{Y}_{CR}\right) \cong \left(RV\left(\widehat{X}_{st}\right) - Cov\left(\widehat{X}_{st}, \widehat{Y}_{st}\right)\right)/X$$
$$= \sum_{i=1}^{K}\left(RV\left(\widehat{X}_i\right) - Cov\left(\widehat{X}_i, \widehat{Y}_i\right)\right)/X$$
$$M\left(\widehat{Y}_{CR}\right) \cong V\left(\widehat{Y}_{st}\right) - 2RCov\left(\widehat{X}_{st}, \widehat{Y}_{st}\right) + R^2V\left(\widehat{X}_{st}\right)$$
$$= \sum_{i=1}^{K}\left(V\left(\widehat{Y}_i\right) - 2RCov\left(\widehat{X}_i, \widehat{Y}_i\right) + R^2V\left(\widehat{X}_i\right)\right)$$
$$\widehat{B}\left(\widehat{Y}_{CR}\right) = \sum_{i=1}^{K}\left(\widehat{R}_C\widehat{V}\left(\widehat{X}_i\right) - \widehat{C}ov\left(\widehat{X}_i, \widehat{Y}_i\right)\right)\Big/\widehat{X},$$

and

$$\widehat{M}\left(\widehat{Y}_{CR}\right) = \sum_{i=1}^{K}\left(\widehat{V}\left(\widehat{Y}_i\right) - 2\widehat{R}_C\widehat{C}ov\left(\widehat{X}_i, \widehat{Y}_i\right) + \widehat{R}_C^2\widehat{V}\left(\widehat{X}_i\right)\right).$$

### 8.7.3 Comparison Between the Separate and Combined Ratio Estimators

(i) To use a separate ratio estimator, one needs to know the values of each stratum's total $X_i$, whereas for the combined ratio estimator, only the knowledge of the population total $X$ is required.

(ii) With regard to bias, the combined ratio estimator should be preferred because the separate ratio estimator is likely to have more bias than the combined one because $X_i$'s are much smaller than $X$. Further, if $R_i = R$ $\forall\, i = 1,\ldots,N$ and each of the stratum biases is of the same sign, then the difference of the absolute biases is

$$\left|B\left(\widehat{Y}_{SR}\right)\right| - \left|B\left(\widehat{Y}_{CR}\right)\right| = \sum_{i=1}^{K}\left|RV\left(\widehat{X}_i\right) - Cov\left(\widehat{X}_i, \widehat{Y}_i\right)\right|\left(\frac{1}{X_i} - \frac{1}{X}\right) \geq 0.$$

(iii) For comparing efficiency, we consider the difference

$$M\left(\widehat{Y}_{CR}\right) - M\left(\widehat{Y}_{SR}\right) \cong \sum_{i=1}^{K}(R_i - R)^2V\left(\widehat{X}_i\right)$$
$$+2\sum_{i=1}^{K}(R_i - R)Cov\left(\widehat{X}_i, \widehat{Y}_i - R\widehat{X}_i\right) \tag{8.7.4}$$

The second term of Eq. (8.7.4) is expected to be small compared to the first term if the regression of $y$ on $x$ in each stratum is linear and passes through the origin. The difference $M\left(\widehat{Y}_{CR}\right) - M\left(\widehat{Y}_{SR}\right)$ is likely to be positive unless $R_i$'s are not equal to $R$, for every $\forall i = 1,\ldots, N$. Hence, a separate ratio estimator is expected to be more efficient than the combined one unless regression of $y$ on $x$ passes through the origin for every stratum.

(iv) The expressions of bias and MSE of $\widehat{Y}_{SR}$ are valid if $n_i$'s are large for all of the strata, which may not hold in reality. But for $\widehat{Y}_{CR}$, the expressions of bias and MSE are valid because the total sample size is expected to be large in all practical situations. In conclusion, one should use a separate ratio estimator if the sample sizes for every stratum is large. The combined ratio estimator is preferable if $R_i$'s are the same for all the strata.

### Example 8.7.1

A factory has three categories of staff: skilled, semiskilled, and unskilled of numbering 50, 75, and 100, respectively. Samples of sizes 10, 12, and 15 are selected from each of the categories by SRSWOR method and information on daily wages and years of service was collected (Table 8.7.1). Estimate the average wages of the factory by using (i) separate and (ii) combined ratio estimators, when the average years of service of the three categories are known to be 15, 8, and 3 years, respectively. Estimate the gain in efficiencies of the separate ratio estimator with respect to the combined ratio estimator.

**Table 8.7.1** Wages and years of service of factory workers

| **Skilled** | | **Semiskilled** | | **Unskilled** | |
|---|---|---|---|---|---|
| **Years of service (years)** | **Wages ($)** | **Year of service (years)** | **Wages ($)** | **Year of service (years)** | **Wages ($)** |
| 10 | 500 | 10 | 450 | 2 | 100 |
| 20 | 600 | 6 | 250 | 4 | 200 |
| 15 | 550 | 12 | 500 | 2 | 150 |
| 25 | 750 | 5 | 350 | 4 | 200 |
| 30 | 800 | 5 | 400 | 6 | 300 |
| 8 | 450 | 12 | 500 | 4 | 250 |
| 5 | 500 | 10 | 450 | 2 | 150 |
| 15 | 650 | 8 | 450 | 5 | 350 |
| 20 | 700 | 10 | 400 | 5 | 350 |
| 15 | 600 | 10 | 550 | 2 | 150 |
| | | 12 | 450 | 4 | 300 |
| | | 8 | 350 | 2 | 250 |
| | | | | 5 | 300 |
| | | | | 4 | 300 |
| | | | | 6 | 550 |

Here we take years of service ($x$) as an auxiliary variable and daily wage as a study variable ($y$). From the table above, we compute the following:

| **Strata** | $N_i$ | $\overline{X}_i$ | $n_i$ | $\overline{x}(s_i)$ | $\overline{y}(s_i)$ | $\widehat{R}_i$ | $N_i\overline{x}(s_i)$ | $N_i\overline{y}(s_i)$ | $\widehat{Y}_{iR}$ | $s^2_{\hat{d}_i}$ | $\frac{W_i^2(1-f_i)s^2_{\hat{d}_i}}{n_i}$ | $s^2_{\hat{d}^*_i}$ | $\frac{W_i^2(1-f_i)s^2_{\hat{d}^*_i}}{n_i}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Skilled | 50 | 15 | 10 | 16.3 | 610 | 37.423 | 815 | 30,500 | 28,067.25 | 34,316.266 | 135.570 | 67,343.138 | 266.046 |
| Semiskilled | 75 | 8 | 12 | 9 | 425 | 47.222 | 675 | 31,875 | 28,333.20 | 7,425.645 | 57.755 | 7,438.108 | 57.851 |
| Unskilled | 100 | 3 | 15 | 3.8 | 260 | 68.421 | 380 | 26,000 | 20,526.30 | 4,360.902 | 48.813 | 4,728.685 | 52.930 |
| Total | 225 | | | | | | 1,870 | 88,375 | 76,926.75 | | 242.138 | | 376.827 |

Separate ratio estimate for the population mean $\overline{Y}$ is $\widehat{\overline{Y}}_{SR} = \sum_{i=1}^{K} \widehat{Y}_{iR} / N = 76,926.75/225 = \$341.90.$

Estimated MSE of $\widehat{\overline{Y}}_{SR}$ is

$$\widehat{M}\left(\widehat{\overline{Y}}_{SR}\right) \sum_{i=1}^{K} \left( \widehat{V}\left(\widehat{Y}_i\right) - 2\widehat{R}_i C\widehat{o}v\left(\widehat{Y}_i, \widehat{X}_i\right) + R_i^2 \widehat{V}\left(\widehat{X}_i\right)\right) / N^2$$

$$= \sum_{i=1}^{K} W_i^2 (1 - f_i) s_{\hat{d}_i}^2 \Big/ n_i = 242.138$$

where $W_i = N_i/N$.

The combined ratio estimator for $\overline{Y}$ is $\widehat{\overline{Y}}_{CR} = \widehat{Y}_{CR}/N = \dfrac{\hat{Y}_{st}}{\hat{X}_{st}} \overline{X}$

$$= \frac{88,375}{1870} \times \frac{1650}{225} = \$346.57.$$

Estimated MSE of $\widehat{\overline{Y}}_{CR}$ is

$$\widehat{M}\left(\widehat{\overline{Y}}_{CR}\right) = \sum_{i=1}^{K} \left( \widehat{V}\left(\widehat{Y}_i\right) - 2\widehat{R}_C Cov\left(\widehat{X}_i, \widehat{Y}_i\right) + \widehat{R}_C V \widehat{X}_i\right) / N^2$$

$$= \sum_{i=1}^{K} W_i^2 (1 - f_i) s_{\hat{d}_i^*}^2 \Big/ n_i$$

$$\left( \text{Where } s_{\hat{d}_i^*}^2 = \sum_{j \in s_i} \left( \widehat{d}_{ij}^* - \overline{\widehat{d}}^* \right)^2 \Big/ (n_i - 1), \widehat{d}_{ij}^* = y_{ij} - \widehat{R}_C x_{ij}, \widehat{R}_C \right.$$

$$\left. = \left(\widehat{Y}_{st}/\widehat{X}_{st}\right) = 47.259 \text{ and } \overline{\widehat{d}}^* = \frac{1}{n_i} \sum_{j \in s_i} \widehat{d}_{ij}^* \right)$$

$$= 376.827$$

Gain in efficiency of the separate ratio estimator over the combined ratio estimator is $\left( \dfrac{\hat{M}\left(\hat{\overline{Y}}_{CR}\right)}{\hat{M}\left(\hat{\overline{Y}}_{SR}\right)} - 1 \right) \times 100 = (376.827/242.138 - 1)$ $\times 100 = 55.62\%$

## 8.8 RATIO ESTIMATOR FOR SEVERAL AUXILIARY VARIABLES

Let us consider the situation where the study variable $y$ is approximately proportional to each of the $p$ auxiliary variables $x_1,\ldots, x_p$. The ratio estimator for the total $Y$ based on the $i$th auxiliary variable $x_i$ when its total $X_i$ is known is given by

$$\widehat{Y}_{R_i} = \frac{\widehat{Y}}{\widehat{X}_i} X_i \quad \text{for } i = 1, \ldots, p \tag{8.8.1}$$

Olkin (1958) proposed the following composite estimator for $Y$ based on the auxiliary variables $x_1,\ldots, x_p$ as

$$\widehat{Y}_{RO} = \sum_{i=1}^{p} w_i \widehat{Y}_{R_i} = \sum_{i=1}^{p} w_i \left( \frac{\widehat{Y}}{\widehat{X}_i} X_i \right) \tag{8.8.2}$$

where $w_i$'s are suitably chosen weights to minimize the MSE of $\widehat{Y}_{RO}$ subject to $\sum_{i=1}^{p} w_i = 1$.

For large sample size $n$, the estimator $\widehat{Y}_{RO}$ is approximately unbiased for $Y$. Hence the MSE of $\widehat{Y}_{RO}$ is approximately equal to its variance and we have

$$\begin{aligned} M\left(\widehat{Y}_{RO}\right) \cong V\left(\widehat{Y}_{RO}\right) &= \sum_{i=1}^{p} w_i^2 V\left(\widehat{Y}_{R_i}\right) + \sum_{i\neq}^{p} \sum_{j=1}^{p} w_i w_j Cov\left(\widehat{Y}_{R_i}, \widehat{Y}_{R_j}\right) \\ &= \sum_{i=1}^{p} w_i^2 C_{ii} + \sum_{i\neq}^{p} \sum_{j=1}^{p} w_i w_j C_{ij} \end{aligned} \tag{8.8.3}$$

where $C_{ii} = V\left(\widehat{Y}_{R_i}\right)$ and $C_{ij} = Cov\left(\widehat{Y}_{R_i}, \widehat{Y}_{R_j}\right)$.

Now minimizing $M\left(\widehat{Y}_{RO}\right)$ subject to $\sum_{i=1}^{p} w_i = 1$, the optimum value of $w_i$ is obtained as

$$w_{i0} = \Delta_i/\Delta \tag{8.8.4}$$

where $\Delta_i$ is the sum of the elements of the $i$th row of the matrix of $\mathbf{C}^{-1}$, $\mathbf{C}$ is the $p \times p$ matrix with $i$th and $j$th element as $C_{ij}$ and $\Delta$ is the sum of all the elements of $\mathbf{C}^{-1}$. The value of $M\left(\widehat{Y}_{RO}\right)$ with $w = w_{i0} = \Delta_i/\Delta$ is denoted by $M_{\min}\left(\widehat{Y}_{RO}\right)$ and is equal to $1/\Delta$.

For example, for $p = 2$, we have $C = \begin{pmatrix} C_{11} & C_{12} \\ C_{12} & C_{22} \end{pmatrix}$,

$\mathbf{C}^{-1} = \dfrac{1}{\left(C_{11}C_{22} - C_{12}^2\right)}\begin{pmatrix} C_{22} & -C_{12} \\ -C_{12} & C_{11} \end{pmatrix}$, $\mathbf{\Delta}_1 = \dfrac{C_{22} - C_{12}}{\left(C_{11}C_{22} - C_{12}^2\right)}$,

$\mathbf{\Delta}_2 = \dfrac{C_{11} - C_{12}}{\left(C_{11}C_{22} - C_{12}^2\right)}$, and $\mathbf{\Delta} = \dfrac{C_{11} + C_{22} - 2C_{12}}{\left(C_{12}C_{22} - C_{12}^2\right)}$. Hence,

$w_{10} = \dfrac{C_{22} - C_{12}}{(C_{11} + C_{22} - 2C_{12})}$, $w_{20} = \dfrac{C_{11} - C_{12}}{(C_{11} + C_{22} - 2C_{12})}$, and

$$M_{\min} = \mathbf{\Delta}^{-1} = \frac{C_{11}C_{22} - C_{12}^2}{C_{11} + C_{22} - 2C_{12}} \tag{8.8.5}$$

Since, $C_{ij}$'s are unknown, the weights $w_{10}$ and $w_{20}$ are estimated by replacing $C_{ij}$ by their suitable estimates.

### 8.8.1 Simple Random Sampling Without Replacement

Let a sample $s$ of size $n$ be selected by the SRSWOR method and $\overline{x}_1, \overline{x}_2$, and $\overline{y}$ be the sample means of the auxiliary variables $x_1$, $x_2$, and the study variable $y$, respectively. Then,

$$\widehat{Y}_{R_1} = \frac{\overline{y}}{\overline{x}_1}X_1,\ \widehat{Y}_{R_2} = \frac{\overline{y}}{\overline{x}_2}X_2,\ \widehat{Y}_{RO} = w_1\widehat{Y}_{R_1} + w_2\widehat{Y}_{R_2},$$

$$C_{11} = V\left(\widehat{Y}_{R_1}\right) \cong \frac{(1-f)}{n}Y^2\left(C_y^2 - 2\rho_1 C_{x_1}C_y + C_{x_1}^2\right)$$

$$C_{22} = V\left(\widehat{Y}_{R_2}\right) \cong \frac{(1-f)}{n}Y^2\left(C_y^2 - 2\rho_2 C_{x_2}C_y + C_{x_2}^2\right),$$

and

$$C_{12} \cong \frac{(1-f)}{n}Y^2\left(C_y^2 - \rho_1 C_{x_1}C_y - \rho_2 C_{x_2}C_y + \rho^* C_{x_1}C_{x_2}\right)$$

where $C_y$ and $C_{xi}$ are CVs of $y$ and $x_i$, $\rho_i$ is the correlation coefficient between $x_i$ and $y$ for $i = 1, 2$, and $\rho^* =$ correlation coefficient between $x_1$ and $x_2$.

In particular if $C_{x1} = C_{x2} = C_x$ and $\rho_1 = \rho_2 = \rho$, we get $C_{11} = C_{22}$ and

$$M_{\min} = \frac{C_{11} + C_{12}}{2}$$

$$= \frac{(1-f)}{n}Y^2\left(C_y^2 - 2\rho C_x C_y + \frac{(1+\rho^*)}{2}C_x^2\right) \tag{8.8.6}$$

$$= V\left(N\overline{y}_s\right) - \frac{(1-f)}{n}Y^2\left(2\rho C_x C_y - \frac{(1+\rho^*)}{2}C_x^2\right)$$

The ratio estimator $\widehat{Y}_{RO}$ will be more efficient than the conventional estimator $\widehat{Y} = N\overline{y}_s$.

If $M_{\min} < V\left(N\overline{y}_s\right)$

$$\text{i.e., if } \left(\frac{\rho}{1+\rho^*}\right)\frac{C_y}{C_x} > \frac{1}{4} \tag{8.8.7}$$

## 8.9 EXERCISES

**8.9.1** Describe the ratio method of estimation. Derive approximate expressions of bias and MSEs under SRSWOR sampling. Derive the conditions of superiority of the ratio estimator over the sample mean for estimating the population mean.

**8.9.2** Consider the following estimators with $\alpha$ as a constant.

(i) $\widehat{Y}_1 = \alpha\widehat{Y}$ (Searls, 1964)

(ii) $\widehat{Y}_2 = \widehat{Y}\left(\dfrac{X}{\hat{X}}\right)^{\alpha}$ (Srivastava, 1967)

(iii) $\widehat{Y}_3 = \alpha\widehat{Y} + (1-\alpha)\widehat{Y}\left(\dfrac{X}{\hat{X}}\right)$ (Chakrabarty, 1968)

(iv) $\widehat{Y}_4 = \alpha\widehat{Y}\left(\dfrac{X}{\hat{X}}\right)$

For each of the estimators above, determine (a) approximate expression of bias, (b) MSE, (c) optimum value of $\alpha$, which minimizes the MSE, and (d) expressions of biases and mean squares with the optimum value of $\alpha$ under SRSWOR sampling.

**8.9.3** Let $s$ be an SRSWOR sample of size $n$ and $t = \overline{y}(s)\dfrac{\overline{x}(\overline{s})}{\overline{X}}$ be an estimator of $\overline{Y}$ with $\overline{x}(\overline{s}) = \sum_{i \notin s} x_i/(N-n)$. Then show that

(i) the bias of $t = S_{xy}/\left(N\overline{X}\right)$,

(ii) $MSE(t) = (1-f)\left(S_y^2 - 2\rho hRS_xS_y + h^2R^2S_x^2\right)\Big/n$, where $h = n/(N-n)$ and $1/N^2 \cong 0$, and (iii) the estimator $t$ is more efficient than the convention ratio estimator $\widehat{\overline{Y}}_R = \overline{y}(s)\dfrac{\overline{x}(s)}{\overline{X}}$ if $n < N/2$ and $0 < \rho\dfrac{cv(y)}{cv(x)} < N\Big/\{2(N-n)\}$ (Srivenkataramana, 1980).

**8.9.4** To determine the access to treatment of HIV/AIDS patients in a certain locality, a sample of 15 clinics from a total of 50 clinics was selected by the SRSWOR method.

| Serial number of clinics | Number of HIV/AIDS patients | Total number of patients |
|---|---|---|
| 1 | 105 | 1000 |
| 2 | 30 | 400 |
| 3 | 65 | 500 |
| 4 | 120 | 1500 |
| 5 | 75 | 1000 |
| 6 | 90 | 850 |
| 7 | 55 | 600 |
| 8 | 90 | 1200 |
| 9 | 45 | 500 |
| 10 | 35 | 300 |
| 11 | 50 | 750 |
| 12 | 80 | 1000 |
| 13 | 120 | 1000 |
| 14 | 90 | 750 |
| 15 | 60 | 500 |

(i) Estimate the proportion of HIV/AIDS patients treated and its standard error.

(ii) Using the ratio method, estimate the total number of HIV/AIDS patients in the locality when the average number of patients treated per clinic is known to be 700. Obtain a 90% confidence interval of the total number of HIV/AIDS patients treated in the locality. Estimate the efficiency of the ratio estimator with the sample mean as an estimator.

**8.9.5** A simple random sample of 600 households in a region from a total of 12,000 households was selected, and the information of household size ($x$) and household income ($y$) was collected. The following is the summarization of data: $\overline{x} = 5.3, \overline{y} = 5000.00, s_x^2 = 0.567, s_y^2 = 150.506$, and $r_{xy} = 0.75$. Estimate the average income of the households and its 95% confidence interval when the average family size is known to be 5, from the past survey.

**8.9.6** A sample of 10 households from a population of 50 households is selected by the PPSWR method of sampling using household size as a size variable. The information of household income and household expenditure on food was obtained and has been presented in the following table:

| Sampled households | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| Size | 5 | 4 | 3 | 2 | 4 | 5 | 6 | 4 | 4 | 3 |
| Expenditure on food ($) | 2000 | 2500 | 1500 | 1000 | 2500 | 3000 | 4000 | 2000 | 1500 | 1200 |
| Income ($) | 5500 | 4500 | 5000 | 2500 | 7500 | 6000 | 8500 | 6000 | 2500 | 5000 |

From the past survey, it is known that the average family size and average income of the population are 4.2 and $5000, respectively. Use this information to estimate the following:

(i) Average expenditure on food by the ratio method of estimation.
(ii) Population proportion of expenditure on food.
(iii) Standard error of the estimators used in (i) and (ii).

**8.9.7** The academics of an university are classified into three categories. From each of the categories, samples are selected by the SRSWOR method, and information on age and earnings is given in the following table.

| | | | Sample | | | | |
|---|---|---|---|---|---|---|---|
| **Categories** | **Number of employees** | **Average age** | **Size** | **Mean ages (years)** | **Mean income ($)** | **SD of ages (years)** | **SD of income ($)** |
| Professor | 50 | 55 | 10 | 60 | 22,000 | 15 | 200 |
| Senior Lecturer | 100 | 40 | 15 | 45 | 15,000 | 25 | 325 |
| Lecturer | 200 | 35 | 30 | 35 | 10,000 | 10 | 250 |

(i) Estimate the average income of the academics of the university by the separate and combined ratio methods of estimation.
(ii) Estimate the relative efficiency of the separate ratio estimator with respect to the combined ratio estimator and comment on your findings.