UNIT-IV

RATIO ESTIMATORS

NEED

Generally the agencies conducting the sample surveys are interested in estimating Population total, means and Proportions But in practice knowledge of ratio of two variables is varying from unit to unit is often required.

For example, one may wish to estimate the ratio of "housing loans to total loans", "Ratio of unemployed persons" "Sex Ratios", "birth rate" etc., To estimate such ratios, commonly used procedure is to take the ratio of unbiased estimator of the numerator (y-variable under study) and denominator (x-auxillary variate) of the population ratio as an estimate and such estimator is known as Ratio Estimator.

The Ratio estimators are precise only when there is an linear regression between the variables and the regression line passing through the origin.

Auxillary Variate

An auxillary variate is one which is highly correlated with the variable under study. Auxillary variate is used to increase the precision of the estimator. Generally in ratio estimation the value of the auxillary variate (x) is often the value of main variate (y) at some previous time.

Ex: To estimate the avg weight of a student, we can make use of the height of the student

Notations :

Let y_i = Value of i^{th} unit of the characteristic under study

x_i = Value of auxillary Variate corresponding to y_i

Y = Population total of y

X = Popwlation total of x

$R = \frac{Y}{X} = \frac{\bar{Y}}{\bar{X}}$ = Population ratio

ρ = correlation coefficient between x & y

Ratio ESTIMATOR - DEFINITION

Suppose that $\bar{y}$ and $\bar{x}$ are the sample means of the characteristics y and x respectively based on a sample of size n selected from N units using SRSWOR

The Ratio estimator of population ratio (R) is given by $\hat{R} = \frac{y}{x} = \frac{\bar{y}}{\bar{x}}$ where $y = \sum_{i=1}^{n} y_i$, $x = \sum_{i=1}^{n} x_i$

To find the ratio estimators for the population mean ($\bar{Y}$) and population total Y, the population total of the variate x say X must be known. The ratio estimator of the population total and population mean is given by

$$\hat{Y}_R = \hat{R} X$$

$$\hat{\bar{Y}}_R = \hat{R} \bar{X}$$

BIAS OF THE RATIO ESTIMATOR

The difference between the avg value of a statistic and a parameter is known as bias involved in estimating the parameter (ie)

$$\text{Bias} = B(t) = E(t) - \theta$$

In ratio estimation, the bias of the ratio estimator is given by $B(\hat{R}) = E(\hat{R}) - R$

Generally the ratio estimators are biased. The ratio estimators are consistent for moderately large samples and the bias is negligible when the sample size is large

Theorem

In simple random sampling the bias of the ratio estimator $\hat{R}$ is

$$B(\hat{R}) = \frac{-\operatorname{cov}(\hat{R}, \bar{x})}{\bar{X}}$$

Proof :

Consider $\operatorname{cov}(\hat{R}, \bar{x}) = E(\hat{R}\bar{x}) - E(\hat{R})E(\bar{x})$

$\because \operatorname{cov}(x, y) = E(xy) - E(x)E(y)$

$$= E\left[\frac{\bar{y}}{\bar{x}} \times \bar{x}\right] - E(\hat{R})\,E(\bar{x})$$

$$= E(\bar{y}) - E(\hat{R})\,E(\bar{x})$$

Since we are using SRSWOR we have

$E(\bar{y}) = \bar{Y}, \quad E(\bar{x}) = \bar{X}$

$\therefore \operatorname{cov}(\hat{R}, \bar{x}) = \bar{Y} - E(\hat{R}) \cdot \bar{X}$

Divide both sides by $\bar{X}$ we have

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

$$= R - E(\hat{R})$$

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

$$B(\hat{R}) = \frac{-\operatorname{cov}(\hat{R}, \bar{x})}{\bar{X}}$$ Hence the proof.

COROLLARY :

Obtain the bias of $\hat{Y}_R$

Proof :

$$B(\hat{Y}_R) = E(\hat{Y}_R) - Y$$

$$= E(\hat{R}X) - Y$$

$$= E(\hat{R}X) - RX$$

$$= X[E(\hat{R}) - R]$$

$$B(\hat{Y}_R) = X\, B(\hat{R})$$

Obtain the bias of $\hat{\bar{Y}}_R$

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

$$= E(\hat{R}\bar{x}) - \bar{Y}$$

$$= E(\hat{R}\bar{x}) - R\bar{x}$$

$$= \bar{x}[E(\hat{R}) - R]$$

$$B(\hat{\bar{Y}}_R) = \bar{x}\, B(\hat{R})$$

Show that the relative bias in $\hat{R}$ ≤ coefficient of variation of $\bar{x}$

Proof :

consider $$B(\hat{R}) = \frac{-\operatorname{cov}(\hat{R}, \bar{x})}{\bar{x}}$$

$$= \frac{-\rho\, \sigma_{\hat{R}}\, \sigma_{\bar{x}}}{\bar{x}}$$

$$\rho = \frac{\operatorname{cov}(x, y)}{\sigma_x \cdot \sigma_y}$$

$$|B(\hat{R})| = \frac{|-\rho\, \sigma_{\hat{R}}\, \sigma_{\bar{x}}|}{\bar{x}}$$

$$\frac{|B(\hat{R})|}{\sigma_{\hat{R}}} = \frac{\rho\, \sigma_{\bar{x}}}{\bar{X}}$$

Since the correlation $\hat{R}$ and $\bar{x}$ cannot exceed one we have

$$\frac{|B(\hat{R})|}{\sigma_{\hat{R}}} \leq \frac{\sigma_{\bar{x}}}{\bar{X}} = CV(\bar{x})$$

Note

when $\rho = 0 \Rightarrow B(\hat{R}) = 0$

Show that to the first approximation, the bias in $\hat{R}$ is given by

$$B(\hat{R}) = \frac{1-f}{n\bar{X}^2}\left[R S_x^2 - \rho S_x S_y\right]$$

$$= \left(\frac{1-f}{n}\right) R \left[C_{xx} - \rho C_x C_y\right]$$

where $C_{xx} = \left(\frac{S_x}{\bar{X}}\right)^2, \quad C_x = \frac{S_x}{\bar{X}}, \quad C_y = \frac{S_y}{\bar{Y}}$

Proof

Consider $\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R$

$$= \frac{\bar{y} - R\bar{x}}{\bar{x}}$$

Add and subtract $\bar{X}$ in denominator

$$\hat{R} - R = \frac{\bar{y} - R\bar{x}}{\bar{X} + (\bar{x} - \bar{X})}$$

$$= \frac{\bar{y} - R\bar{x}}{\bar{X}\left[1 + \frac{\bar{x} - \bar{X}}{\bar{X}}\right]}$$

$$= \frac{\bar{y} - R\bar{x}}{\bar{X}}\left[1 + \frac{\bar{x} - \bar{X}}{\bar{X}}\right]^{-1}$$

Apply Taylor's Series expansion we get

$$= \frac{\bar{y} - R\bar{x}}{\bar{X}} \left[1 - \left(\frac{\bar{x} - \bar{X}}{\bar{X}} \right) + \left(\frac{\bar{x} - \bar{X}}{\bar{X}} \right)^2 - \cdots \right]$$

Ignoring the terms of Second and higher Order of $\frac{\bar{x} - \bar{X}}{\bar{X}}$

we get $\hat{R} - R = \frac{\bar{y} - R\bar{x}}{\bar{X}} \left[1 - \left(\frac{\bar{x} - \bar{X}}{\bar{X}} \right) \right]$

Taking Expectation on both sides, we have

$$E[\hat{R} - R] = \frac{1}{\bar{X}} E \left\{ (\bar{y} - R\bar{x}) \left[1 - \left(\frac{\bar{x} - \bar{X}}{\bar{X}} \right) \right] \right\}$$

$$B(\hat{R}) = \frac{1}{\bar{X}} \left\{ E(\bar{y} - R\bar{x}) - E\left[(\bar{y} - R\bar{x}) \left(\frac{\bar{x} - \bar{X}}{\bar{X}} \right) \right] \right\}$$

$$= \frac{1}{\bar{X}} \left\{ E(\bar{y} - R\bar{x}) - E\left[\frac{(\bar{y} - R\bar{x})(\bar{x} - \bar{X})}{\bar{X}} \right] \right\} \rightarrow (1)$$

Consider

$$E(\bar{y} - R\bar{x}) = E(\bar{y}) - RE(\bar{x})$$

$$= \bar{Y} - R\bar{X}$$

$$= \bar{Y} - \bar{Y} = 0 \rightarrow (2)$$

Consider

$$E\left[(\bar{y} - R\bar{x})(\bar{x} - \bar{X}) \right] = E\left[\bar{y}(\bar{x} - \bar{X}) - R\bar{x}(\bar{x} - \bar{X}) \right]$$

Add & subtract $\bar{Y}(\bar{x} - \bar{X})$ we get

$$E\left[(\bar{y} - R\bar{x})(\bar{x} - \bar{X}) \right] = E\left[\bar{y}(\bar{x} - \bar{X}) - \bar{Y}(\bar{x} - \bar{X}) + \bar{Y}(\bar{x} - \bar{X}) - R\bar{x}(\bar{x} - \bar{X}) \right]$$

$$= E\left[(\bar{y} - \bar{Y})(\bar{x} - \bar{X}) + R\bar{X}(\bar{x} - \bar{X}) - R\bar{x}(\bar{x} - \bar{X}) \right]$$

$$= E\left[(\bar{y} - \bar{Y})(\bar{x} - \bar{X}) - R(\bar{x} - \bar{X})(\bar{x} - \bar{X}) \right]$$

$$= \operatorname{cov}(\bar{x}, \bar{y}) - R \cdot V(\bar{x}) \rightarrow (3)$$

Since we are using SRSWOR, we have

$\text{Cov}(\bar{x}, \bar{y}) = \frac{1-f}{n} S_{xy}$ and $V(\bar{x}) = \left(\frac{1-f}{n}\right) S_x^2 \longrightarrow (4)$

using (4), (3) can be written as

$$E\left[(\bar{y} - R\bar{x})(\bar{x} - \bar{X})\right] = \left(\frac{1-f}{n}\right) S_{xy} - R\left(\frac{1-f}{n}\right) S_x^2$$

$$= \frac{1-f}{n}\left[S_{xy} - RS_x^2\right] \longrightarrow (5)$$

Using (2) & (5), eqn (1) can be written as

$$B(\hat{R}) = \frac{1}{\bar{X}}\left[0 - \left(\frac{1-f}{n\bar{X}}\right)\left(S_{xy} - RS_x^2\right)\right]$$

$$B(\hat{R}) = \frac{1-f}{n\bar{X}^2}\left[RS_x^2 - S_{xy}\right]$$

$$= \frac{1-f}{n\bar{X}^2} R\left[S_x^2 - \frac{S_{xy}}{R}\right]$$

$$= \frac{1-f}{n\bar{X}^2} R\left[S_x^2 - \frac{\bar{X}}{\bar{Y}} \rho S_x S_y\right] \qquad \because R = \frac{\bar{Y}}{\bar{X}}, \quad \rho = \frac{S_{xy}}{S_x S_y}$$

$$= \frac{1-f}{n} R\left[\frac{S_x^2}{\bar{X}^2} - \rho \frac{S_y}{\bar{Y}} \frac{S_x}{\bar{X}}\right]$$

$$= \frac{1-f}{n} R\left[\left(\frac{S_x}{\bar{X}}\right)^2 - \rho\left(\frac{S_y}{\bar{Y}}\right)\left(\frac{S_x}{\bar{X}}\right)\right]$$

$$B(\hat{R}) = \frac{1-f}{n} R\left[C_{xx} - \rho C_x C_y\right]$$

Hence the proof.

COROLLARY :

Obtain an expression for $B(\hat{Y}_R)$

Soln :

$$B(\hat{Y}_R) = X B(\hat{R})$$

$$= X \frac{1-f}{n\bar{x}^2}\left[R S_x^2 - S_{xy}\right]$$

$$= N\bar{x} \frac{1-f}{n\bar{x}^2}\left[R S_x^2 - S_{xy}\right] \quad \because X = N\bar{X}$$

$$= N\left(\frac{1-f}{n\bar{x}}\right)\left[R S_x^2 - S_{xy}\right]$$

Obtain an expression for $B(\hat{Y}_R)$

$$B(\hat{Y}_R) = B(\bar{X}\hat{R})$$

$$= \bar{X} B(\hat{R})$$

$$= \bar{X} \frac{1-f}{n\bar{X}^2}\left[R S_x^2 - S_{xy}\right]$$

$$= \frac{1-f}{n\bar{X}}\left[R S_x^2 - S_{xy}\right]$$

Theorem :

In SRSWOR, for larger n, an approximation to $V(\hat{R})$ is given by

$$V(\hat{R}) = \frac{1-f}{n\bar{X}^2} \frac{1}{N-1} \sum_{i=1}^{N} (y_i - R x_i)^2$$

Proof

Consider, $$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R$$

$$= \frac{\bar{y} - \bar{x}R}{\bar{x}}$$

$$= \frac{\bar{y} - \bar{x}R}{\bar{X}} \quad \because \bar{x} \simeq \bar{X}$$

Squaring on both sides, we get

$$(\hat{R} - R)^2 = \left(\frac{\bar{y} - \bar{x}R}{\bar{X}}\right)^2$$

Taking expectation on both sides

$$E(\hat{R} - R)^2 = \frac{E(\bar{y} - R\bar{x})^2}{\bar{X}^2}$$

$$V(\hat{R}) = \frac{E(\bar{y} - R\bar{x})^2}{\bar{X}^2} \longrightarrow (1)$$

Consider the variate $u_i = y_i - Rx_i$

$$\bar{u} = \bar{y} - R\bar{x}$$

$$\bar{U} = \bar{Y} - R\bar{X} = 0 \qquad \because \bar{Y} = R\bar{X}$$

Consider,

$$V(\bar{u}) = E(\bar{u} - \bar{U})^2$$

$$= E(\bar{u})^2 \qquad \because \bar{U} = 0$$

$$= E(\bar{y} - R\bar{x})^2 \longrightarrow (2)$$

Consider

$$V(\bar{u}) = \left(\frac{1-f}{n}\right) S_u^2$$

$$= \left(\frac{1-f}{n}\right) \frac{1}{N-1} \sum_{i=1}^{N} (u_i - \bar{U})^2 \qquad \because S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$$

$$= \left(\frac{1-f}{n}\right) \frac{1}{N-1} \sum_{i=1}^{N} u_i^2 \qquad \because \bar{U} = 0$$

$$= \left(\frac{1-f}{n}\right) \frac{1}{N-1} \sum_{i=1}^{N} (y_i - Rx_i)^2 \longrightarrow (3)$$

Using (2) & (3), eqn (1) can be written as

$$V(\hat{R}) = \frac{V(\bar{u})}{\bar{X}^2}$$

$$= \left(\frac{1-f}{n\bar{X}^2}\right) \frac{1}{N-1} \sum_{i=1}^{N} (y_i - Rx_i)^2$$

Hence the proof.

Note :

An unbiased estimator $V(\hat{R})$ is given by

$$V(\hat{R}) = \frac{1-f}{n(n-1)\bar{x}^2} \sum_{i=1}^{n} (y_i - \hat{R}x_i)^2$$

COROLLARY

Obtain $V(\hat{Y}_R)$ and its unbiased estimator

Proof :

Consider $V(\hat{Y}_R) = V(X\hat{R})$

$$= X^2 V(\hat{R})$$

$$= X^2 \frac{1-f}{n\bar{X}^2} \frac{1}{N-1} \sum_{i=1}^{N} (y_i - Rx_i)^2$$

An unbiased estimate of $V(\hat{Y}_R)$ is

$$v(\hat{Y}_R) = \cancel{V(X\hat{R})}\ X^2 \frac{1-f}{n(n-1)\bar{X}^2} \left[\sum_{i=1}^{n} y_i^2 + \hat{R}^2 \sum_{i=1}^{n} x_i^2 - 2\hat{R} \sum_{i=1}^{n} x_i y_i \right]$$

$$= \cancel{\bar{X}^2} N^2 \frac{1-f}{n(n-1)\cancel{\bar{X}^2}} \left[\Sigma y_i^2 + \hat{R}^2 \Sigma x_i^2 - 2\hat{R} \Sigma x_i y_i \right] \quad \because X = N\bar{X}$$

$$= N^2 \frac{1-f}{n(n-1)} \left[\Sigma y_i^2 + \hat{R}^2 \Sigma x_i^2 - 2\hat{R} \Sigma x_i y_i \right]$$

Obtain $V(\hat{\bar{Y}}_R)$ and its unbiased

$$V(\hat{\bar{Y}}_R) = V(\bar{X}\hat{R})$$

$$= \bar{X}^2 V(\hat{R})$$

$$= \cancel{\bar{X}^2} \left(\frac{1-f}{n\cancel{\bar{X}^2}} \right) \frac{1}{N-1} \sum_{i=1}^{N} (y_i - Rx_i)^2$$

$$v(\hat{\bar{Y}}_R) = \frac{1-f}{n(n-1)} \left[\Sigma y_i^2 + \hat{R}^2 \Sigma x_i^2 - 2\hat{R} \Sigma x_i y_i \right]$$

Theorem

Show that to the first order approximation $V(\hat{R})$ is given by

$$V(\hat{R}) = \frac{1-f}{n\bar{X}^2} \left[S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y \right]$$

$$= \left(\frac{1-f}{n} \right) R^2 \left[C_{yy} + C_{xx} - 2\rho C_x C_y \right]$$

Proof :

consider

$$V(\hat{R}) = \frac{1-f}{n\bar{x}^2} \frac{1}{N-1} \sum_{i=1}^{N} (y_i - R x_i)^2$$

Add and subtract $\bar{Y}$

$$V(\hat{R}) = \frac{1-f}{n\bar{x}^2} \frac{1}{N-1} \sum_{i=1}^{N} [y_i - \bar{y} + \bar{y} - R x_i]^2$$

$$= \frac{1-f}{n\bar{x}^2} \frac{1}{N-1} \sum_{i=1}^{N} [y_i - \bar{y} + R\bar{x} - R x_i]^2 \quad \because \bar{y} = R\bar{x}$$

$$= \frac{1-f}{n\bar{x}^2} \frac{1}{N-1} \sum_{i=1}^{N} [(y_i - \bar{y}) - R(x_i - \bar{x})]^2$$

$$= \frac{1-f}{n\bar{x}^2} \frac{1}{N-1} \left[\sum_{i=1}^{N} (y_i - \bar{y})^2 + R^2 \sum_{i=1}^{N} (x_i - \bar{x})^2 - 2R \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \right]$$

$$= \frac{1-f}{n\bar{x}^2} \left[\frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N-1} + R^2 \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1} - \frac{2R\sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y})}{N-1} \right]$$

$$= \frac{1-f}{n\bar{x}^2} \left[S_y^2 + R^2 S_x^2 - 2R S_{xy} \right] \quad \because R = Y$$

$$= \left(\frac{1-f}{n}\right) R^2 \left[\frac{S_y^2}{R^2 \bar{x}^2} + \frac{S_x^2}{\bar{x}^2} - \frac{2 S_{xy}}{R\bar{x}^2} \right]$$

$$= \left(\frac{1-f}{n}\right) R^2 \left[\frac{\bar{x}^2 S_y^2}{\bar{y}^2 \bar{x}^2} + \frac{S_x^2}{\bar{x}^2} - \frac{2\bar{x} \rho S_x S_y}{\bar{y}\bar{x}^2} \right]$$

$$\because R = \bar{Y}/\bar{x}, \quad \rho = \frac{S_{xy}}{S_x S_y}$$

$$= \left(\frac{1-f}{n}\right) R^2 \left[\left(\frac{S_y}{\bar{y}}\right)^2 + \left(\frac{S_x}{\bar{x}}\right)^2 - 2\rho \left(\frac{S_x}{\bar{x}}\right)\left(\frac{S_y}{\bar{y}}\right) \right]$$

$$= \left(\frac{1-f}{n}\right) R^2 \left[C_{yy} + C_{xx} - 2 \rho C_x C_y \right]$$

$$V(\hat{R}) = \left(\frac{1-f}{n}\right) R^2 \left[C_{yy} + C_{xx} - 2 C_{xy} \right] \quad \because \rho = \frac{C_{xy}}{C_x C_y}$$

Hence the proof.

Comparison of Ratio estimator with mean per unit.

For large sample using SRS, the ratio estimator $(\hat{Y}_R)$ has smaller variance than SRS estimator if

$$\rho > \frac{1}{2} \left[\frac{S_x / \bar{x}}{S_y / \bar{y}} \right] = \frac{1}{2} \frac{C_x}{C_y}$$

Proof :

In SRSWOR the variance of the estimator $\hat{Y} = N\bar{y}$ is given by

$$V(\hat{Y}) = N^2 \left(\frac{1-f}{n}\right) S_y^2 \longrightarrow (1)$$

The variance of the ratio estimator of the population total $\hat{Y}_R = \hat{R} X$ is given by

$$V(\hat{Y}_R) = N^2 \left(\frac{1-f}{n}\right) \left[S_y^2 + R^2 S_x^2 - 2 R \rho S_x S_y \right] \longrightarrow (2)$$

Consider,

$$V(\hat{Y}_R) - V(\hat{Y}) = N^2 \left(\frac{1-f}{n}\right) \left[S_y^2 + R^2 S_x^2 - 2 R \rho S_x S_y \right] - N^2 \left(\frac{1-f}{n}\right) S_y^2$$

$$= N^2 \left(\frac{1-f}{n}\right) \left[R^2 S_x^2 - 2 R \rho S_x S_y \right] < 0$$

The ratio estimator $(\hat{Y}_R)$ has smaller variance than SRS estimator $\hat{Y}$

If $V(\hat{Y}_R) - V(\hat{Y}) < 0$

$$\Rightarrow N^2 \left(\frac{1-f}{n}\right) \left[R^2 S_x^2 - 2 R \rho S_x S_y \right] < 0$$

$$R^2 S_x^2 - 2R\rho S_x S_y < 0$$

$$R S_x \left[R S_x - 2 \rho S_y \right] < 0$$

$$R S_x - 2 \rho S_y < 0$$

$$R S_x < 2 \rho S_y$$

$$\rho > \frac{1}{2} R \frac{S_x}{S_y}$$

$$= \frac{1}{2} \left(\frac{S_x / \bar{X}}{S_y / \bar{Y}} \right) \qquad \because R = \bar{Y}/\bar{X}$$

$$= \frac{1}{2} \frac{C_x}{C_y}$$

REGRESSION ESTIMATOR

Like Ratio estimators, linear regression estimators also make use of the auxillary information for increasing the precision of the estimate. Ratio estimators is best among the class of estimators only when the regression between y and x is linear and the line passing through the origin. when regression is linear and does not passing through the origin, it is better to use regression estimators based on linear regression.

Let us suppose that the simple random sample of size n has been drawn from a population of N units and let the sample means $\bar{x}$ and $\bar{y}$ are calculated. Also the population mean of the variate x say $\bar{X}$ must be known.

The linear regression estimator for the population mean $\bar{Y}$ is given by

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$$

where b is the regression coefficient of y on x and it is

Computed from the sample.

The linear regression estimator for the population total (Y) is given by

$$\hat{Y}_{lr} = N \cdot \bar{y}_{lr}$$

Note :

when $b=0$, $\bar{y}_{lr} = \bar{y} + 0(\bar{X} - \bar{x})$

$= \bar{y}$ = Mean is SRS

when $b=\hat{R}$, $\bar{y}_{lr} = \bar{y} + \hat{R}(\bar{X} - \bar{x})$

$= \bar{y} + \hat{R}\bar{X} - \hat{R}\bar{x}$ $\quad\because \hat{R} = \frac{\bar{y}}{\bar{x}}$

$= \bar{y} + \hat{R}\bar{X} - \bar{y}$

$= \hat{R}\bar{X}$

$\therefore \bar{y}_{lr} = \hat{\bar{Y}}_R$

Theorem :

Show that $\bar{y}_{lr}$ is an unbiased estimate of the population mean 'b' being a fixed quantity say $b = b_0$

Proof :

Consider $\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$

It is given that b being a fixed quantity say $b = b_0$

$$\bar{y}_{lr} = \bar{y} + b_0(\bar{X} - \bar{x})$$

Taking expectation on both sides

$$E(\bar{y}_{lr}) = E(\bar{y}) + b_0 E(\bar{X} - \bar{x})$$

$$= E(\bar{y}) + b_0[\bar{X} - E(\bar{x})]$$

Since we use SRS, we have $E(\bar{y}) = \bar{Y}$ and $E(\bar{x}) = \bar{X}$

$$\therefore E(\bar{y}_{lr}) = \bar{Y} + b_0(\bar{X} - \bar{X})$$

$$\Rightarrow E(\bar{y}_{lr}) = \bar{Y}$$ Hence the proof.

Note :

An unbiased estimate of the population total is given by $\hat{Y}_{lr} = N\bar{y}_{lr}$

Theorem :

In SRS of size n with pre-assigned $b = b_0$, show that

$$V(\bar{y}_{lr}) = \frac{1-f}{n}\left[S_y^2 + b_0^2 S_x^2 - 2b_0 \rho S_x S_y\right]$$

and $V(\bar{y}_{lr})_{min} = \left(\frac{1-f}{n}\right)(1-\rho^2) S_y^2$

Proof :

Consider,

$$V(\bar{y}_{lr}) = E\left[\bar{y}_{lr} - E(\bar{y}_{lr})\right]^2$$

$$= E(\bar{y}_{lr} - \bar{Y})^2 \qquad \because E(\bar{y}_{lr}) = \bar{Y}$$

$$= E\left[\bar{y} + b_0(\bar{X} - \bar{x}) - \bar{Y}\right]^2$$

$$= E\left[(\bar{y} - \bar{Y}) + b_0(\bar{X} - \bar{x})\right]^2$$

$$= E\left[(\bar{y} - \bar{Y}) - b_0(\bar{x} - \bar{X})\right]^2$$

$$V(\bar{y}_{lr}) = E\left[(\bar{y} - \bar{Y})^2 + b_0^2(\bar{x} - \bar{X})^2 - 2b_0(\bar{x} - \bar{X})(\bar{y} - \bar{Y})\right]$$

$$= E(\bar{y} - \bar{Y})^2 + b_0^2 E(\bar{x} - \bar{X})^2 - 2b_0 E\left[(\bar{x} - \bar{X})(\bar{y} - \bar{Y})\right]$$

$$= V(\bar{y}) + b_0^2 V(\bar{x}) - 2b_0 \, cov(\bar{x}, \bar{y})$$

$$= \frac{1-f}{n} S_y^2 + b_0^2 \frac{1-f}{n} S_x^2 - 2b_0\left(\frac{1-f}{n}\right) S_{xy}$$

$\because$ we use SRS for selecting sample

$$= \frac{1-f}{n}\left[S_y^2 + b_0^2 S_x^2 - 2b_0 S_{xy}\right]$$

$$= \frac{1-f}{n}\left[S_y^2 + b_0^2 S_x^2 - 2b_0 \rho S_x S_y\right] \longrightarrow (1)$$

$$\because \rho = \frac{S_{xy}}{S_x S_y}$$

We have to find the value of b_0 which minimizes $V(\bar{y}_{lr})$

(ie) $V(\bar{y}_{lr})$ is minimum only when $\frac{\partial V(\bar{y}_{lr})}{\partial b_0} = 0$ and

$$\frac{\partial^2 V(\bar{y}_{lr})}{\partial b_0^2} > 0$$

Consider,

$$\frac{\partial V(\bar{y}_{lr})}{\partial b_0} = \frac{1-f}{n}\left[0 + 2b_0 S_x^2 - 2\rho S_x S_y\right]$$

$$\frac{\partial V(\bar{y}_{lr})}{\partial b_0} = 0 \Rightarrow \left(\frac{1-f}{n}\right)\left[2 b_0 S_x^2 - 2\rho S_x S_y\right] = 0$$

$$\Rightarrow 2 b_0 S_x^2 - 2\rho S_x S_y = 0$$

$$2 S_x \left[b_0 S_x - \rho S_y\right] = 0$$

$$b_0 S_x - \rho S_y = 0$$

$$b_0 S_x = \rho S_y$$

$$b_0 = \frac{\rho S_y}{S_x} = B$$

where B = population regression coefficient of y on x.

Also $$\frac{\partial^2 V(\bar{y}_{lr})}{\partial b_0^2} = \left(\frac{1-f}{n}\right) 2 S_x^2 > 0$$

$$= 2\left(\frac{1-f}{n}\right) S_x^2 > 0$$

The minimum value of $V(\bar{y}_{lr})$ is obtained by substituting

$b_0 = \frac{\rho S_y}{S_x}$ in (1)

$$V(\bar{y}_{lr})_{min} = \left(\frac{1-f}{n}\right)\left[S_y^2 + \rho^2 \frac{S_y^2}{\cancel{S_x^2}} \cancel{S_x^2} - 2\rho \frac{S_y}{\cancel{S_x}} \rho \cancel{S_x} S_y\right]$$

$$= \left(\frac{1-f}{n}\right)\left[S_y^2 + \rho^2 S_y^2 - 2\rho^2 S_y^2\right]$$

$$= \frac{1-f}{n}\left[S_y^2 - \rho^2 S_y^2 \right]$$

$$= \left(\frac{1-f}{n}\right) S_y^2 (1-\rho^2)$$

Note :

An unbiased estimator of $V(\bar{y}_{lr})$ is

$$v(\bar{y}_{lr}) = \frac{1-f}{n}\left[s_y^2 + b_0^2 s_x^2 - 2 b_0 s_{xy} \right]$$

An estimate of the population total is given by $\hat{Y} = N[\bar{X} + (\bar{y} - \bar{x})]$. Show that it is unbiased and obtain its standard error.

Soln :

Consider $\hat{Y} = N[\bar{X} + (\bar{y} - \bar{x})]$

Taking expectations on both sides

$$E(\hat{Y}) = N E[\bar{X} + (\bar{y} - \bar{x})]$$

$$= N[E(\bar{X}) + E(\bar{y}) - E(\bar{x})]$$

$$= N[\bar{X} + \bar{Y} - \bar{X}]$$

$$= N\bar{Y} = Y$$

ie) $\hat{Y}$ is an unbiased estimate of Y.

consider, $$V(\hat{Y}) = V\left[N[\bar{X} + (\bar{y} - \bar{x})]\right]$$

$$= N^2 V[\bar{X} + (\bar{y} - \bar{x})]$$

$$= N^2 \left[V(\bar{X}) + V(\bar{y} - \bar{x}) - 2\,cov[\bar{X}, (\bar{y} - \bar{x})]\right] \rightarrow (1)$$

$\because V(x_1 + x_2) = V(x_1) + V(x_2) - 2\,cov(x_1, x_2)$

$V(\bar{X}) = 0 \rightarrow (2)$ $\quad \because \bar{X}$ is constant, $V(\text{const}) = 0$

$$V(\bar{y}-\bar{x}) = V(\bar{y}) + V(\bar{x}) - 2\,cov(\bar{y},\bar{x})$$

$$= \left(\frac{1-f}{n}\right) S_y^2 + \left(\frac{1-f}{n}\right) S_x^2 - 2\left(\frac{1-f}{n}\right) S_{xy} \longrightarrow (3)$$

Consider,

$$cov(\bar{x},(\bar{y}-\bar{x})) = E\left[\bar{x}(\bar{y}-\bar{x})\right] - E(\bar{x})\,E(\bar{y}-\bar{x})$$

$$= \bar{X}\,E(\bar{y}-\bar{x}) - \bar{X}\,E(\bar{y}-\bar{x})$$

$$= 0 \longrightarrow (4)$$

Using (2), (3) and (4) eqn (1) can be written as

$$V(\hat{Y}) = N^2\left[0 + \left(\frac{1-f}{n}\right) S_y^2 + \left(\frac{1-f}{n}\right) S_x^2 - 2\left(\frac{1-f}{n}\right) S_{xy}\right]$$

$$= N^2\left(\frac{1-f}{n}\right)\left[S_y^2 + S_x^2 - 2S_{xy}\right]$$

$$S.E(\hat{Y}) = \sqrt{V(\hat{Y})} = N\sqrt{\frac{1-f}{n}}\sqrt{S_y^2 + S_x^2 - 2S_{xy}}$$

(or)

$$V(\hat{Y}) = E\left[\hat{Y} - E(\hat{Y})\right]^2$$

$$= E\left[\hat{Y} - Y\right]^2$$

$$= E\left\{N\left[\bar{x} + (\bar{y}-\bar{x})\right] - Y\right\}^2$$

$$= E\left[N\left[\bar{x} + (\bar{y}-\bar{x})\right] - N\bar{Y}\right]^2 \qquad \bar{Y} = \frac{Y}{N}$$

$$= E\left[N\bar{x} - N\bar{Y} + N(\bar{y}-\bar{x})\right]$$

$$= E\left[N\left[(\bar{x}-\bar{Y}) + (\bar{y}-\bar{x})\right]\right]^2$$

$$= N^2 E\left[(\bar{x}-\bar{Y}) + (\bar{y}-\bar{x})\right]^2$$

$$V(\hat{Y}) = N^2 E\left[(\bar{y}-\bar{Y}) - (\bar{x}-\bar{X})\right]^2$$

$$= N^2\left[E(\bar{y}-\bar{Y})^2 + E(\bar{x}-\bar{X})^2 - 2E(\bar{x}-\bar{X})(\bar{y}-\bar{Y})\right]$$

$= N^2 \left[V(\bar{y}) + V(\bar{x}) - 2\,cov(\bar{x}, \bar{y}) \right]$

$= N^2 \left(\frac{1-f}{n} \right) \left[S_y^2 + S_x^2 - 2 S_{xy} \right]$

BIAS IN REGRESSION ESTIMATOR

In SRS, the bias of $\bar{y}_{lr}$ is approximately is given by

$B(\bar{y}_{lr}) \simeq -cov(\bar{x}, b)$

Proof:

Consider, the relative error in sample mean $\bar{y}$ as

$$e = \frac{\bar{y} - \bar{Y}}{\bar{Y}}$$

Relative error : $e = \frac{t - \theta}{\theta}$

$\bar{Y} e = \bar{y} - \bar{Y}$

$\bar{y} = \bar{Y} + \bar{Y} e$

$= \bar{Y}(1+e) \rightarrow (1)$

Similarly $e_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}$

$\bar{x} = \bar{X}(1+e_1) \rightarrow (2)$

$e_2 = \frac{b - B}{B}$

$b = B(1+e_2) \longrightarrow (3)$

Now, $E(e) = \frac{E(\bar{y} - \bar{Y})}{\bar{Y}} = \frac{E(\bar{y}) - \bar{Y}}{\bar{Y}}$

$= \frac{\bar{Y} - \bar{Y}}{\bar{Y}} = 0$

Hly $E(e_1) = 0$ and $E(e_2) = 0$

we have

$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$

using (1), (2) and (3) we can write $\bar{y}_{lr}$ as

$$\bar{y}_{lr} = \bar{Y}(1+e) + B(1+e_2)\left[\bar{X} - \bar{X}(1+e_1)\right]$$

$$= \bar{Y} + \bar{Y}e + (B + Be_2)\left[\cancel{\bar{X}} - \cancel{\bar{X}} - \bar{X}e_1\right]$$

$$= \bar{Y} + \bar{Y}e + (B + Be_2)(-\bar{X}e_1)$$

$$\bar{y}_{lr} = \bar{Y} + \bar{Y}e - B\bar{X}e_1 - B\bar{X}e_1e_2$$

Taking expectation on both sides

$$E(\bar{y}_{lr}) = \bar{Y} + \bar{Y}E(e) - B\bar{X}E(e_1) - B\bar{X}E(e_1e_2)$$

$$= \bar{Y} + \bar{Y}(0) - B\bar{X}(0) - B\bar{X}E(e_1e_2)$$

$$= \bar{Y} - B\bar{X}E(e_1e_2)$$

$$E(\bar{y}_{lr}) - \bar{Y} = -B\bar{X}E(e_1e_2)$$

$\because E(\bar{y}_{lr}) - \bar{Y} = B(\bar{y}_{lr})$

$$B(\bar{y}_{lr}) = -B\bar{X}E\left[\left(\frac{\bar{x}-\bar{X}}{\bar{X}}\right)\left(\frac{b-B}{B}\right)\right]$$

$$= \frac{-\cancel{B}\cancel{\bar{X}}E\left[(\bar{x}-\bar{X})(b-B)\right]}{\cancel{\bar{X}}\cancel{B}}$$

$$= -E\left[(\bar{x}-\bar{X})(b-B)\right]$$

$$\therefore B(\bar{y}_{lr}) = -\operatorname{cov}(\bar{x}, b)$$

Note:

Bias of $\hat{Y}_{lr}$ is given by

$$B(\hat{Y}_{lr}) = B(N\bar{y}_{lr})$$

$$= N\,B(\bar{y}_{lr})$$

Large sample comparison of Ratio Regression and SRSWOR Estimates:

For these comparisons the sample size n must be large enough so that the approximate formulae of

Variances of ratio and regression estimators are valid

Consider $V(\bar{y})_{SRS} = \left(\frac{1-f}{n}\right) S_y^2 \rightarrow (1)$

$$V(\hat{\bar{Y}}_R) = \left(\frac{1-f}{n}\right)\left[S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y\right] \rightarrow (2)$$

$$V(\bar{y}_{lr})_{min} = \left(\frac{1-f}{n}\right) S_y^2 (1-\rho^2) \longrightarrow (3)$$

Comparing (2) and (3), it is clear that the regression estimate is always better than SRSWOR estimate only when $\rho \neq 0$.

The regression estimate is precise than the ratio estimate if $V(\bar{y}_R) - V(\bar{y}_{lr}) > 0$

(ie) $\left(\frac{1-f}{n}\right)\left[S_y^2 + R^2 S_x^2 - 2R\rho S_x S_y\right] - \left(\frac{1-f}{n}\right) S_y^2 (1-\rho^2) > 0$

$$\left(\frac{1-f}{n}\right)\left[\cancel{S_y^2} + R^2 S_x^2 - 2R\rho S_x S_y - \cancel{S_y^2} + \rho^2 S_y^2\right] > 0$$

$$[R S_x - \rho S_y]^2 > 0$$

$$S_x \left[R - \rho \frac{S_y}{S_x}\right]^2 > 0$$

$$\because B = \frac{\rho S_y}{S_x}$$

$$[R - B]^2 > 0$$

Thus the regression estimate is more precise than the ratio estimate only when $B \neq R$. This occurs when the regression between y and x is linear and passing through the origin.

Regression Estimate when b is computed from the Sample :

when b is computed from the sample, it is similar estimate of B

to that of the least square

i.e) $b = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$

In this case the linear regression estimate of the population mean is $\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$

Relationship between b and B :

Let us define the variate

$$e_i = (y_i - \bar{Y}) - B(x_i - \bar{X})$$

$$\sum_{i=1}^{N} e_i = \sum_{i=1}^{N} \left[(y_i - \bar{Y}) - B(x_i - \bar{X}) \right]$$

$$= \sum_{i=1}^{N} y_i - N\bar{Y} - B \sum_{i=1}^{N} x_i + BN\bar{X}$$

$$= N\bar{Y} - N\bar{Y} - BN\bar{X} + BN\bar{X} = 0$$

consider

$$\sum_{i=1}^{N} e_i (x_i - \bar{X}) = \sum_{i=1}^{N} \left[(y_i - \bar{Y}) - B(x_i - \bar{X}) \right] (x_i - \bar{X})$$

$$= \sum_{i=1}^{N} (y_i - \bar{Y})(x_i - \bar{X}) - B \sum_{i=1}^{N} (x_i - \bar{X})^2$$

$$= B \sum_{i=1}^{N} (x_i - \bar{X})^2 - B \sum_{i=1}^{N} (x_i - \bar{X})^2 = 0$$

$\rho = \dfrac{S_{xy}}{S_x S_y}$

$\dfrac{\rho S_y}{S_x} = B$

$B = \dfrac{S_{xy}}{S_x^2}$

$B = \dfrac{\sum (x_i - \bar{X})(y_i - \bar{Y})}{\sum (x_i - \bar{X})^2}$

Consider,

$$b = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x}) - \bar{y} \sum_{i=1}^{n} (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} y_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \qquad \because \sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

$$= \frac{\sum_{i=1}^{n} \left[e_i + \bar{Y} + B(x_i - \bar{x}) \right] (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} e_i (x_i - \bar{x}) + \bar{Y} \sum_{i=1}^{n} (x_i - \bar{x}) + B \sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{\sum_{i=1}^{n} e_i (x_i - \bar{x}) + B (x_i - \bar{x})^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$\because \sum_{i=1}^{n} (x_i - \bar{x}) = 0$

$\bar{x} \simeq \bar{X}$

$$b = B + \frac{\sum_{i=1}^{n} e_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$b - B = \frac{\sum_{i=1}^{n} e_i (x_i - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Note that $b - B$ is of order $\frac{1}{\sqrt{n}}$

Theorem :

If b is the least square estimate of B, then in Simple random sampling of size n

$$V(\bar{y}_{lr}) = \left(\frac{1-f}{n}\right) S_y^2 (1 - \rho^2)$$

Provided n is large enough so that the terms of order $\frac{1}{\sqrt{n}}$ are negligible.

Proof

Consider

$$e_i = (y_i - \bar{Y}) - B(x_i - \bar{X})$$

By averaging e_i over all the units in the sample we have,

$$\bar{e} = \frac{\sum_{i=1}^{n} e_i}{n} = \frac{1}{n}\sum_{i=1}^{n}\left[(y_i - \bar{Y}) - B(x_i - \bar{X})\right]$$

$$= \frac{\sum_{i=1}^{n} y_i}{n} - \frac{n\bar{Y}}{n} - B\frac{\sum_{i=1}^{n} x_i}{n} + \frac{Bn\bar{X}}{n}$$

$$= \bar{y} - \bar{Y} - B\bar{x} + B\bar{X}$$

$$\bar{e} = (\bar{y} - \bar{Y}) - B(\bar{x} - \bar{X}) \rightarrow (1)$$

Consider,

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$$

$$= \bar{e} + \bar{Y} + B(\bar{x} - \bar{X}) + b(\bar{X} - \bar{x}) \quad \text{using (1)}$$

$$= \bar{e} + \bar{Y} - B(\bar{X} - \bar{x}) + b(\bar{X} - \bar{x})$$

$$= \bar{e} + \bar{Y} + (b - B)(\bar{X} - \bar{x}) \rightarrow (2)$$

$(b-B)$ is of order $\frac{1}{\sqrt{n}}$ and also $(\bar{X} - \bar{x})$ is also of order $\frac{1}{\sqrt{n}}$

$(b-B)(\bar{X} - \bar{x})$ is of order $\frac{1}{n}$.

Hence the product $(b-B)(\bar{X}-\bar{x})$ is negligible, we ignore

Since the terms of order $1/\sqrt{n}$ is

the product term.

$\therefore$ eqn (2) becomes

$$\bar{y}_{lr} = \bar{e} + \bar{Y}$$

$$\bar{e} = \bar{y}_{lr} - \bar{Y}$$

Consider

$$E(\bar{e}) = E[\bar{y}_{lr} - \bar{Y}]$$

$$= \bar{Y} - \bar{Y} = 0$$

$$\therefore V(\bar{e}) = E[\bar{e} - E(\bar{e})]^2$$

$$= E(\bar{e})^2$$

$$= E\left[\bar{y}_{lr} - \bar{Y}\right]^2$$

$$= V(\bar{y}_{lr})$$

(ie) $V(\bar{y}_{lr}) = V(\bar{e}) = \left(\frac{1-f}{n}\right) S_e^2 \longrightarrow (3)$

Consider,

$$\bar{e}_N = \frac{\sum_{i=1}^{N} e_i}{N}$$

$$= \frac{1}{N}\left[\sum y_i - N\bar{Y} - B\sum x_i + BN\bar{X}\right]$$

$$= \frac{1}{N}\left[N\bar{Y} - N\bar{Y} - BN\bar{X} + BN\bar{X}\right]$$

$$= 0$$

$$S_e^2 = \frac{1}{N-1}\sum_{i=1}^{N}(e_i - \bar{e}_N)^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N} e_i^2$$

$$= \frac{1}{N-1}\sum_{i=1}^{N}\left[(y_i - \bar{Y}) - B(x_i - \bar{X})\right]^2$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2 + B^2\sum_{i=1}^{N}(x_i-\bar{X})^2 - 2B\sum_{i=1}^{N}(x_i-\bar{X})(y_i-\bar{Y})\right]$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2 + B^2\sum_{i=1}^{N}(x_i-\bar{X})^2 - 2B^2\sum_{i=1}^{N}(x_i-\bar{X})\right]$$

$$\because B = \frac{\sum(x_i-\bar{X})(y_i-\bar{Y})}{\sum(x_i-\bar{X})^2}$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N}(y_i-\bar{Y})^2 - B^2\sum_{i=1}^{N}(x_i-\bar{X})^2\right]$$

$$= \frac{\sum_{i=1}^{N}(y_i-\bar{Y})^2}{N-1} - B^2\frac{\sum_{i=1}^{N}(x_i-\bar{X})^2}{N-1}$$

$$= S_y^2 - B^2 S_x^2$$

$$\because B = \frac{S_{xy}}{S_x^2} = \frac{S_{xy}}{S_x \cdot S_x}$$

$$B S_x = \frac{S_{xy}}{S_x}$$

$$\rho = \frac{S_{xy}}{S_x \cdot S_y}$$

$$\rho S_y = \frac{S_{xy}}{S_x}$$

$$\Rightarrow B S_x = \rho S_y$$

$$S_e^2 = S_y^2 - \rho^2 S_y^2$$

$$S_e^2 = S_y^2(1-\rho^2) \longrightarrow (4)$$

Using (4), eqn (3) becomes

$$V(\bar{y}_{lr}) = \left(\frac{1-f}{n}\right) S_y^2 (1-\rho^2)$$

In SRSWOR, for the estimator $\hat{\bar{Y}} = \bar{y} - \lambda(\bar{x} - \bar{X})$, find the mean and variance of $\hat{\bar{Y}}$, where λ is a constant. What value of λ, $V(\hat{\bar{Y}})$ is minimum and also obtain the simple estimator (SRS), Ratio and Regression estimators as a particular cases of $\hat{\bar{Y}}$.

Proof :

$$\hat{\bar{Y}} = \bar{y} - \lambda(\bar{x} - \bar{X})$$

$$\hat{\bar{Y}} = \bar{y} + \lambda(\bar{X} - \bar{x})$$

Taking expectations on both sides

$$E(\hat{\bar{Y}}) = E(\bar{y}) + \lambda E(\bar{X} - \bar{x})$$

$$= \bar{Y} + \lambda[\bar{X} - E(\bar{x})]$$

$$= \bar{Y} + \lambda[\bar{X} - \bar{X}]$$

$$= \bar{Y}$$

Consider

$$V(\hat{\bar{Y}}) = E[\hat{\bar{Y}} - E(\hat{\bar{Y}})]^2$$

$$= E[\hat{\bar{Y}} - \bar{Y}]^2$$

$$= E[\bar{y} + \lambda(\bar{X} - \bar{x}) - \bar{Y}]^2$$

$$= E[(\bar{y} - \bar{Y}) - \lambda(\bar{x} - \bar{X})]^2$$

$$= E(\bar{y} - \bar{Y})^2 + \lambda^2 E(\bar{x} - \bar{X})^2 - 2\lambda E[(\bar{x} - \bar{X})(\bar{y} - \bar{Y})]$$

$$V(\hat{\bar{Y}}) = V(\bar{y}) + \lambda^2 V(\bar{x}) - 2\lambda \, cov(\bar{x}, \bar{y})$$

$$= \left(\frac{1-f}{n}\right) S_y^2 + \lambda^2 \left(\frac{1-f}{n}\right) S_x^2 - 2\lambda \left(\frac{1-f}{n}\right) S_{xy}$$

$$= \frac{1-f}{n}\left[S_y^2 + \lambda^2 S_x^2 - 2\lambda S_{xy}\right]$$

$$= \frac{1-f}{n}\left[S_y^2 + \lambda^2 S_x^2 - 2\lambda \rho S_x S_y\right] \longrightarrow (1)$$

we have to find the value of λ which minimize the $V(\hat{\bar{y}})$

$V(\hat{\bar{y}})$ is minimum only when $\frac{\partial V(\hat{\bar{y}})}{\partial \lambda} = 0$ and

$\frac{\partial^2 V(\hat{\bar{y}})}{\partial \lambda^2} > 0$

Consider,

$$\frac{\partial V(\hat{\bar{y}})}{\partial \lambda} = \left(\frac{1-f}{n}\right)\left[0 + 2\lambda s_x^2 - 2\rho s_x s_y\right]$$

$$\frac{\partial V(\hat{\bar{y}})}{\partial \lambda} = 0 \Rightarrow \left(\frac{1-f}{n}\right)\left[2\lambda s_x^2 - 2\rho s_x s_y\right] = 0$$

$$2\lambda s_x^2 - 2\rho s_x s_y = 0$$

$$2 s_x \left[\lambda s_x - \rho s_y\right] = 0$$

$$\lambda s_x - \rho s_y = 0$$

$$\lambda = \frac{\rho s_y}{s_x} = B$$

also $\frac{\partial^2 V(\hat{\bar{y}})}{\partial \lambda^2} = \left(\frac{1-f}{n}\right) 2 s_x^2 > 0$

The minimum value of $V(\hat{\bar{y}})$ is obtained by substituting

$\lambda = \frac{\rho s_y}{s_x}$ in eqn (1)

$$V(\hat{\bar{y}})_{min} = \left(\frac{1-f}{n}\right)\left[s_y^2 + \frac{\rho^2 s_y^2}{s_x^2} s_x^2 - 2\rho \frac{s_y}{s_x} \rho s_x s_y\right]$$

$$= \left(\frac{1-f}{n}\right)\left[s_y^2 + \rho^2 s_y^2 - 2\rho^2 s_y^2\right]$$

$$= \left(\frac{1-f}{n}\right)\left[s_y^2 - \rho^2 s_y^2\right]$$

$$= \left(\frac{1-f}{n}\right) s_y^2 (1-\rho^2)$$

case ii) $\lambda = 0$

$$\hat{\bar{y}} = \bar{y} - 0(\bar{x} - \bar{X}) = \bar{y}$$

i.e) $\hat{\bar{Y}}$ is nothing but the sample mean of SRS

In this case eqn (1) becomes

$$V(\hat{\bar{Y}}) = \left(\frac{1-f}{n}\right) S_y^2$$

case (ii) : $\lambda = \hat{R}$, $\hat{R} = \bar{y}/\bar{x}$

$$\hat{\bar{Y}} = \bar{y} - \frac{\bar{y}}{\bar{x}}(\bar{x} - \bar{X})$$

$$= \bar{y} - \frac{\bar{y}}{\cancel{\bar{x}}}\cancel{\bar{x}} - \frac{\bar{y}}{\bar{x}}\bar{X}$$

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

In this case. eqn (1) becomes

$$V(\hat{\bar{Y}}) = \left(\frac{1-f}{n}\right)\left[S_y^2 + \hat{R}^2 S_x^2 - 2\hat{R}\rho S_x S_y\right]$$

$$= V(\hat{\bar{Y}}_R)$$

Case (iii) $\lambda = b$

$$\cancel{\bar{y}} \quad \hat{\bar{Y}} = \bar{y} - b(\bar{x} - \bar{X})$$

$$= \bar{y} + b(\bar{X} - \bar{x})$$

$$= \bar{y}_{lr}$$

In this case eqn (1) becomes

$$V(\hat{\bar{Y}}) = \left(\frac{1-f}{n}\right)\left[S_y^2 + b^2 S_x^2 - 2b\rho S_x S_y\right]$$

$$\therefore V(\hat{\bar{Y}}) = V(\bar{y}_{lr})$$