UNIT - III

Stratified Random Sampling:-

If a simple random sample of size $n_i, (i=1,2,\ldots k)$ is drawn form each of the stratum respectively such that $n = \sum_{i=1}^{k} n_i$, the sample is termed as stratified Random sample of size n and the technique of drawing such a sample is called stratified Random sampling.

Estimate of population Mean and variance:-

Mean:- y_{st} is an unbiased estimate of the population mean $\bar{Y}_N$, i.e., $E(y_{st}) = \bar{Y}_N \rightarrow (1)$

Proof: Since the sample in each of the stratum is a simple random sample, we have,

$$E(\bar{y}_{ni}) = \bar{Y}_{Ni}$$

$$\therefore E(y_{st}) = \frac{1}{N}\sum_{i=1}^{k} N_i \cdot E(y_{ni}) = \frac{1}{N}\sum_{i=1}^{k} N_i \bar{Y}_{Ni} = \bar{Y}_N.$$

Variance:

$$Var(y_{st}) = \frac{1}{N^2}\sum_{i=1}^{k} N_i(N_i - n_i)\left(\frac{S_i^2}{n_i}\right) = \sum_{i=1}^{k} P_i^2\left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2$$

Proof:- since the sample in each stratum is simple random sample without replacement, we have,

$$Var(y_{ni}) = \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2 \rightarrow ①$$

Where S_i^2 is as defined ;

$$S_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (Y_{ij} - \bar{Y}_{Ni})^2 \quad (i=1,2,\ldots k).$$

$$var(\bar{y}_{st}) = Var\left(\sum_{i=1}^{K} P(y_{ni})\right) = \sum_{i=1}^{K} P_i^2 \cdot var(y_{ni})$$

$$\therefore Var(\bar{y}_{st}) = \sum_{i=1}^{K} P_i^2 \cdot \left(\frac{1}{n_i} - \frac{1}{N_i}\right) S_i^2. \quad [\text{from } ①]$$

$$= \frac{1}{N^2} \sum_{i=1}^{K} N_i (N_i - n_i) \cdot \frac{S_i^2}{n_i}.$$

Optimum Allocation:- Another guiding principle in the determination of the n_i's is to choose them so as to:

(a) Minimise the variance of the estimate for

(i) fixed sample , (ii) fixed cost.

(b) Minimise the total cost for fixed desired precision

The allocation of n_i's to various strata in accordance with the above principles is known as optimum allocation.

Thus in optimum allocation n_i's are to be obtained such that,

(1) $Var(\bar{y}_{st})$ is minimum for mixed n.

(2) $var(\bar{y}_{st})$ is minimum for fixed total cost c.

(3) Total cost c is minimum for fixed value of $Var(\bar{y}_{st}) = V_0$.

The result concerning the above points are given in the following theorems.

Cost function:- In any sample survey the value of information on the experimented units must always be balaned against the cost of obtaining it.

Theorem:

$Var(\bar{y}_{st})$ is minimum for fixed total size of the sample (n) if $n_i \propto N_i S_i$

Proof:- Here the problem is to minimise,

$$var(\bar{y}_{st}) = \frac{1}{N^2} \sum_{i=1}^{k} N_i(N_i - n_i) \frac{S_i^2}{n_i} \rightarrow (1)$$

subject to the condition,

$$\sum_{i=1}^{k} n_i = n$$

finally we get,

$$n_i = \frac{n N_i S_i}{\sum_{i=1}^{k} N_i S_i} \rightarrow (2).$$

Thus in optimum allocation for a fixed total sample size, we have. $n_i \propto N_i S_i \rightarrow (3)$

This is known as Neyman's formula for optimum allocation.

Proportional Allocation: Allocation of n_i's to various strata is called proportional if the sample fraction is constant for each stratum,

$$\frac{n_1}{N_1} = \frac{n_2}{N_2} = \cdots = \frac{n_k}{N_k} = \frac{\Sigma n_i}{\Sigma N_i} = \frac{n}{N} = c \text{ (constant)}.$$

$$\Rightarrow \frac{n_i}{N_i} = c = \frac{n}{N} \Rightarrow n_i \propto N_i, \ (i = 1, 2, \ldots k) \rightarrow (1)$$

In proportional allocation the $var(\bar{y}_{st})$ is given by:

$$Var(\bar{y}_{st})_{prop} = \frac{1}{N^2} \sum_{i=1}^{k} N_i(N_i - n_i) \frac{S_i^2}{n_i}$$

$$= \sum_{i=1}^{k} \frac{P_i}{N} \cdot \left[\frac{N}{n} - 1\right] S_i^2 \quad \left[\because \frac{N_i}{n_i} = \frac{N}{n}\right]$$

$$Var(\bar{y}_{st})_{prop} = \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} P_i S_i^2 \quad \rightarrow (2).$$

Advantages of stratified Random Sampling:-

1. More Representative:- In an unstratified random sample some strata may be over-represented, others may be under-represented while some may be excluded altogether. stratified sampling ensures any desired representation in the sample of the various strata in the population.

2. Greater Accuracy:- stratified sampling provides estimates with increased precision. Moreover, stratified sampling enables us to obtain the results of known precision for each of the stratum.

3. Administrative convenience:- As compared with simple random sample, the stratified samples would be more concentrated geographically. Accordingly, the time and money involved in collecting the data and interviewing the individuals may be considerably reduced and the supervision of the field work could be allocated with greater ease and convenience.

4. Sometimes, the sampling problems may differ markedly in different parts of the population. e.g. a population under study consisting of (i) Literates and illiterate or (ii) people living in institutions, hostels, hospitals, etc, and those living in ordinary homes.

Comparison of stratified Random sampling with sample Random sampling without stratification:-

Proportional Allocation Vs simple Random sampling:-

The variance of the estimate of population mean in stratified random sampling with proportional allocation.

$$Var(\bar{y}_{st})_p = \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{K} p_i s_i^2 \rightarrow (1).$$

and $$var(\bar{y}_n)_R = \left(\frac{1}{n}-\frac{1}{N}\right)\cdot S^2 \rightarrow (2)$$

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{K}\sum_{j=1}^{Ni}(Y_{ij}-\bar{Y}_{Ni}+\bar{Y}_{Ni}-\bar{Y}_N)^2$$

$$= \sum_{i=1}^{K}(Ni-1)S_i^2 + \sum_{i=1}^{K} Ni(\bar{Y}_{Ni}-\bar{Y}_N)^2$$

The product term vanishes since $\sum_{j=1}^{Ni}(y_{ij}-\bar{Y}_{Ni})$, being the algebraic sum of the deviations from mean is zero. If we assume that Ni and consequently N are sufficiently large so that Ni and consequently N are sufficiently large so that we can take $Ni-1 \simeq Ni$ and $N-1 \simeq N$, then we get,

$$NS^2 \simeq \sum_{i=1}^{k} Ni S_i^2 + \sum_{i=1}^{k} Ni(\bar{Y}_{Ni}-\bar{Y}_N)^2$$

$$\simeq Var(\bar{y}_{st})_p + \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{k} P_i(\bar{Y}_{Ni}-\bar{Y}_N)^2 \rightarrow (3)$$

$$\Rightarrow var(\bar{y}_n)_R \geq var(\bar{y}_{st})_p \rightarrow (4).$$

Neyman's Allocation Vs Proportional Allocation:-

Writing $var(\bar{y}_{st})_p$ and $var(\bar{y}_{st})_N$ for the variance of the estimate of population mean in stratified sampling with proportional allocation and Neyman's optimum allocation respectively, we get from, (7.48a) and (7.48a).

$$var(\bar{y}_{st})_p - var(\bar{y}_{st})_N = \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=1}^{k} p_i s_i^2 - \left(\frac{1}{n}(\Sigma p_i s_i)^2 - \frac{1}{N}\Sigma p_i s_i^2\right)$$

$$= \frac{1}{n}\left[\sum_{i=1}^{k} p_i s_i^2 - \left(\sum_{i=1}^{k} p_i s_i\right)^2\right]$$

$$\text{Var}(\bar{y}_{st})_p - \text{Var}(\bar{y}_{st})_N = \frac{1}{n} \sum_{i=1}^{k} p_i (s_i - \bar{s})^2 \rightarrow ①$$

Where $\bar{s} = \sum_{i=1}^{k} p_i s_i = \frac{1}{N} \sum_{i=1}^{k} N_i s_i$ is the weighted mean of the stratum standard deviations, the weights being equal to the stratum sizes.

Since R.H.S in (7.57a) we conclude that Neyman's optimum allocation gives better estimates than proportional allocation and greater the difference between the stratum standard deviations, more is the gain in precision of Neyman's allocation over proportional allocation.

Neyman's Allocation Vs Simple Random sampling:-

Substituting for $\text{var}(\bar{y}_{st})_p$ from ① in ③ we get,

$$\text{Var}(\bar{y}_{st})_R = \text{var}(\bar{y}_{st})_N + \frac{1}{n} \sum_{i=1}^{k} p_i (s_i - \bar{s})^2 + \left(\frac{1}{n} - \frac{1}{N}\right) \sum_{i=1}^{k} p_i (\bar{Y}_{Ni} - \bar{Y}_N)^2$$

$$\text{Var}(\bar{y}_n)_R \geq \text{var}(\bar{y}_{st})_N \rightarrow ②.$$

We observe from ① that as we change from unstratified simple random sampling to stratified random sampling with Neyman's allocations, the gain in precision of the estimates results from two components, viz.,

(i) The elimination of the difference between stratum means and.

(ii) the elimination of the difference among stratum standard deviations.

UNIT - IV

Systematic sampling:-

Systematic sampling consists in drawing a random number, say $i \le k$ and selecting the unit corresponding to this number and every kth unit subsequently. Thus the systematic sample of size n will consist of the units

$$i,\ i+k,\ i+2k,\ i+(n-1)k.$$

Estimation of population mean and variance:-

Mean:

Random Start	Sample composition (units in the sample)	probability	Mean.
1	$1 \quad 1+k \cdots 1+jk \cdots 1+(n-1)k$	$1/k$	$\bar{y}_{1.}$
2	$2 \quad 2+k \cdots 2+jk \cdots 2+(n-1)k$	$1/k$	$\bar{y}_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	$i \quad i+k \cdots i+jk \cdots i+(n-1)k$	$1/k$	$\bar{y}_{i.}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$k \quad 2k \cdots (1+j)k \cdots nk$	$1/k$	$\bar{y}_{k.}$

Thus k rows of the table give the k-systematic samples. The columns of the above table are also sometimes refferred to as n strata.

$$E(\bar{y}_i) = \frac{1}{k}\sum_{i=1}^{k} \bar{y}_i = \bar{y}_{..} \quad \rightarrow (1)$$

$\bar{y}_{sys}$ = mean of the systematic sample $= \bar{y}_i$,

$$\therefore Var(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k} (\bar{y}_{i} - \bar{y}_{..})^2 \quad \rightarrow (2)$$

Variance :-

$$Var(\bar{y}_{sys}) = \frac{N-1}{N} S^2 - \frac{(n-1)k}{N} S^2_{wsy} \rightarrow ①$$

Where,

$$S^2_{wsy} = \frac{1}{k(n-1)} = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_i)^2 \rightarrow ②$$

Proof:-

We have,

$$(N-1) S^2 = \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.} + \bar{y}_{i.} - \bar{y}_{..})^2$$

$$= \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{i.})^2 + \sum_{i=1}^{k} \sum_{j=1}^{n} (\bar{y}_{i.} - \bar{y}_{..})^2$$

the co-variance term vanishes, since,

$$\sum_i \sum_j (y_{ij} - \bar{y}_{i.}) (\bar{y}_{i.}) (\bar{y}_{i.} - \bar{y}_{..}) = \sum_{i=1}^{k} \left[(\bar{y}_{i.} - \bar{y}_{..}) \sum_{i=1}^{n} (y_{ij} - \bar{y}_{i.}) \right] = 0$$

$$\therefore (N-1) S^2 = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2 + n \sum_{i=1}^{k} (y_i - \bar{y}_{..})^2$$

$$= k(n-1) S^2_{wsy} + nk \, var(\bar{y}_{sys}). \quad [\text{From ① and ②}]$$

$$\Rightarrow var(\bar{y}_{sys}) = \frac{N-1}{N} \cdot S^2 - \frac{k(n-1)}{N} S^2_{wsy} \quad [\because N = nk].$$

Merits of systematic sampling:-

(1) Systematic sampling is operationally more convenient than simple random sampling or stratified random sampling. Time and work involved is also relatively much less.

(2) Systematic sampling may be more efficient than simple random sampling provided the frame is arranged wholly at random. The most common approach to

to randomness is provided by alphabetical list such as names in telephone directly, although even these may have certain non-random characteristics.

comparison of SRS:-

Systematic sampling vs simple Random sampling:-

In case of srswor,

$$\text{Var}(\bar{y}_R) = \frac{N-n}{Nn} \cdot S^2$$

where $S^2 = \frac{1}{nk-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (y_{ij} - \bar{y}_{..})^2$.

$$\text{Var}(\bar{y}_n) - \text{var}(\bar{y}_{sys}) = \left(\frac{N-n}{n} - N+1\right)\frac{S^2}{N} + \frac{k(n-1)}{N} \cdot S^2_{wsy}$$

$$= \frac{k(n-1)}{N} \cdot S^2_{wsy} - \frac{n-1}{n} \cdot S^2$$

$$= \frac{n-1}{n}\left(S^2_{wsy} - S^2\right) \rightarrow ① \quad [\because N = nk].$$

The systematic sampling gives more precise estimate of the population mean as compared with srswor if and only if

$$\text{Var}(\bar{y}_n) - \text{Var}(\bar{y}_{sys}) > 0$$

$$\Rightarrow S^2_{wsy} > S^2 \rightarrow ②$$

This leads to the following important conclusion:-

" A systematic sample is more precise than a simple random sample without replacement if the mean square within the systematic sample is larger than the population mean square". In other words systematic sampling will yield better results only if the units within the same sample are heterogeneous.

Systematic sampling Vs stratified Random sampling:-

Mean of the jth stratum is

$$\bar{y}_{\cdot j} = \frac{1}{k}\sum_{i=1}^{k} y_{ij}, \quad (j=1,2,\dots n) \rightarrow ①$$

population mean is,

$$\bar{y}_{..} = \frac{1}{nk}\cdot\sum_i\sum_j y_{ij} = \frac{1}{k}\sum_{i=1}^{k} y_{i\cdot} = \frac{1}{n}\sum_{j=1}^{n}\bar{y}_{\cdot j} \rightarrow ②$$

Stratum mean square is,

$$s_j^2 = \frac{1}{N_j-1}\sum_{i=1}^{k}(y_{ij}-\bar{y}_{\cdot j})^2 = \frac{1}{k-1}\cdot\sum_{i=1}^{k}(y_{ij}-\bar{y}_{\cdot j})^2 \rightarrow ③$$

$[\because N_j = k; j=1,2,\dots n]$.

s_{wst}^2 = pooled mean square between units within strata,

$$s_{wst}^2 = \frac{1}{n(k-1)}\cdot\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\cdot j})^2 \rightarrow ④$$

ρ_{wst} is the correlation co-efficient between deviations from stratum means of pairs of items that are in the same systematic sample. thus,

$$\rho_{wst} = \frac{E(y_{ij}-\bar{y}_{\cdot j})(y_{ij'}-\bar{y}_{\cdot j'})}{E(y_{ij}-\bar{y}_{\cdot j})^2}$$

$$\Rightarrow \frac{1}{kn(n-1)}\cdot\frac{\sum_{i=1}^{k}\sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y}_{\cdot j})(y_{ij'}-\bar{y}_{\cdot j'})}{\frac{1}{nk}\cdot\sum_{i=1}^{k}\sum_{j=1}^{n}(y_{ij}-\bar{y}_{\cdot j})^2}$$

$$= \frac{\sum_{i=1}^{k}\sum_{j\neq j'=1}^{n}(y_{ij}-\bar{y}_{\cdot j})(y_{ij'}-\bar{y}_{\cdot j'})}{(n-1)n(k-1)s_{wst}^2} \rightarrow ⑤$$

$[\because$ from ④$]$

Theorem:-

If the population consists of a linear trend, then prove that,

$$Var(\bar{y}_{st}) \leq Var(\bar{y}_{sys}) \leq (\bar{y}_n)_R.$$

Proof:- Let us suppose that the population has the linear trend given by the model.

$$y_i = i;\ (i = 1, 2, \ldots, N).$$

then,
$$\sum_{i=1}^{N} y_i = \sum_{i=1}^{N} i = \frac{N(N+1)}{2}.$$

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} i^2 = \frac{N(N+1)(2N+1)}{6}.$$

$$\bar{Y}_N = \frac{1}{N} \cdot \sum_{i=1}^{N} y_i = \frac{N+1}{2}.$$

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

$$= \frac{1}{N-1} \left[\frac{N(N+1)(2N+1)}{6} - \frac{N(N+1)^2}{4} \right]$$

[on simplification].

$$\therefore var(\bar{y}_n)_R = \left(\frac{1}{n} - \frac{1}{N}\right) S^2 = \frac{N-n}{Nn} S^2$$

$$= \frac{(k-1)(nk+1)}{12} \rightarrow (1).$$

$$Var(\bar{y}_{st}) = \frac{k-1}{n^2 k} \cdot \sum_{j=1}^{n} s_i^2$$

We have $S^2 = \dfrac{N(N+1)}{12}$

For population of N units, since j^{th} stratum consists of k units, we have $s_i^2 = \dfrac{k(k+1)}{12}.$

$$\therefore \text{Var}(\bar{y}_{st}) = \frac{k-1}{n^2 k} \cdot \frac{nk(k+1)}{12} = \frac{k^2-1}{12n} \rightarrow (2)$$

for finding out var($\bar{y}_{sys}$), we have.

$\bar{y}_{i.}$ = mean of the values of ith sample.

$$= \frac{1}{n}\sum_{j=1}^{n} y_{ij} = \frac{1}{n}\left[i+(i+k)+(i+2k \cdots + i+(n-1)k\right]$$

$$= \frac{1}{n}\left[(ni + \{1+2+\cdots+(n-1)\}k\right]. = i + \frac{(n-1)}{2}k \rightarrow (3)$$

Also. $\bar{y}_{..} = \bar{Y}_N = \frac{N+1}{2} = \frac{nk+1}{2} \rightarrow (4).$

$$\Rightarrow \bar{y}_{i.} - \bar{y}_{..} = i - \frac{k+1}{2}.$$

$$\therefore \text{var}(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}\left[i^2 + \left(\frac{k+1}{2}\right)^2 - \frac{(k+1)}{\cancel{2}}\, \cancel{2}\, i\right] \quad \left[\text{from: } \text{var}(\bar{y}_{sys}) = \frac{1}{k}\sum_{i=1}^{k}(\bar{y}_{i.}-\bar{y}_{..})^2\right]$$

$$= \frac{(k+1)(2k+1)}{6} + \frac{(k+1)^2}{4} - \frac{(k+1)^2}{2}$$

$$= \frac{k^2-1}{12} \quad \text{[on simplification]} \rightarrow (5).$$

From (1), (2) and (5), we get,

$$\text{var}(\bar{y}_{st}) : \text{Var}(\bar{y}_{sys}) : \text{var}(\bar{y}_n) :: \frac{k+1}{n} : (k+1) : (nk+1)$$

$$\simeq \frac{1}{n} : 1 : n \text{ (approx).}$$

$$\Rightarrow \text{Var}(\bar{y}_{st}) \leq \text{Var}(\bar{y}_{sys}) \leq \text{Var}(\bar{y}_n).$$

Thus if the population is subject of a linear trend then stratified random sampling is most effective in eliminating the effect of linear trend.

UNIT - V

Ratio Estimators:-

In the ratio method, an auxillary variable x_i correlated with y_i is obtained for each unit of the population. The population total X of x_i must be known. In practice x_i is often the value of y_i at some previous time, when a complete census was taken. The aim in this method is to obtain increased precision by taking advantage of the corelation co-efficient between x_i and y_i, Atpresent we consider only SRS.

Theorem 1:-

In SRSWOR for large n, an approximation for the variation of $\hat{R}$ is given by,

$$\text{Var}(\hat{Y}_R) = \frac{N^2(1-f)}{n}\left[\sum_{i=1}^{N} \frac{(y_i - Rx_i)^2}{N-1}\right] \rightarrow (1)$$

$$\text{Var}(\bar{Y}_R) = \frac{1-f}{n}\left[\sum_{i=1}^{N} \frac{(y_i - Rx_i)^2}{N-1}\right] \rightarrow (2)$$

$$\text{Var}(\hat{R}) = \frac{1-f}{n\bar{x}^2}\left[\sum_{i=1}^{N} \frac{(y_i - Rx_i)^2}{N-1}\right] \rightarrow (3)$$

Where,

$f = \frac{n}{N}$ is the sampling fraction,

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}} X, \quad \hat{Y}_R \text{ (or) } \hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}} \cdot \bar{X}$$

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

Proof:-

$$\hat{R} - R = \frac{\bar{y}}{\bar{x}} - R. = \frac{\bar{y} - R\bar{x}}{\bar{x}} \rightarrow ④$$

We replace $\bar{x}$ as $\bar{X}$ in the denominator of ④

We get, $$\hat{R} - R = \frac{\bar{y} - R\bar{x}}{\bar{X}} \rightarrow ⑤$$

By definition.

$$V(x) = E[(x - E(x))]^2$$

Here, $$var(\hat{R}) = E[\hat{R} - E(\hat{R})]^2$$

$$\frac{E(\bar{y})}{E(\bar{x})} = \frac{\bar{Y}}{\bar{X}}.$$

$$\therefore E(\hat{R}) = R.$$

$$var(\hat{R}) = E(\hat{R} - R)^2 \rightarrow ⑥.$$

$$var(\bar{d}) = (1-f)\frac{s^2}{n} \rightarrow ⑦$$

$$var(\hat{R}) = \frac{1}{\bar{X}^2} var(\bar{d}) \rightarrow ⑧$$

By applying ⑦ in ⑧ becomes,

$$var(\hat{R}) = \frac{1}{\bar{X}^2}(1-f)\frac{s^2 d}{n} \qquad \left[\because s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2\right].$$

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

$$Var(\hat{R}\bar{X}) = \bar{X}^2 \, var(\hat{R}).$$

$$var(\hat{Y}_R) = Var(N\hat{\bar{Y}}_R) = N^2 Var(\hat{\bar{Y}}_R).$$

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

Theorem : 2

$$Var(\hat{R}) = \frac{(1-f)}{n\bar{x}^2}\left[s^2y + R^2 s^2x - 2\rho R s_y s_x\right].$$

$$Var(\hat{R}) = \frac{(1-f)}{n} R^2 \left[cy^2 + cx^2 - 2\rho cx\, cy\right]$$

Where. $\rho = \sum_{i=1}^{N} \frac{(y_i - \bar{Y})(x_i - \bar{X})}{(N-1) S_x S_y}$

Proof:-

We know that, $R = \frac{\bar{Y}}{\bar{X}}$

$$R\bar{X} = \bar{Y} \rightarrow (1)$$

$$Var(\hat{R}) = \frac{1-f}{n\bar{x}^2} \sum_{i=1}^{N} \frac{(y_i - Rx_i)^2}{N-1} \rightarrow (2).$$

$$\therefore d_i = (y_i - \bar{Y}) - R(x_i - \bar{X}) \rightarrow (3).$$

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

We know that, $\rho = \sum_{i=1}^{N} \frac{(y_i - \bar{Y})(x_i - \bar{X})}{(N-1) S_x S_y}.$

We have,

$$Var(\hat{R}) = \frac{1-f}{n\bar{x}^2}\left[s^2y + R^2 s^2x - 2\rho R S_x S_y\right]$$ [$\because$ since $s^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2 / n-1$].

[$\because$ Dividing throughly by $(N-1)$].

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

$$Var(\hat{R}) = \frac{(1-f)}{n} R^2\left[cy^2 + c_x^2 - 2\rho c_x\, cy\right].$$

Theorem 3:

$$\rho > \frac{1}{2}\left(\frac{S_x}{\bar{X}}\right) = \frac{1 \text{ co-efficient of variation of } x_i}{2 \text{ co-efficient of variation of } y_i}.$$

Proof:-

For $\hat{Y}$ we have,

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

Now,

$$var(\bar{y}) = var(N\bar{y}).$$

$$var(ax) = a^2 \, var(x)$$

$$\Rightarrow N^2 \, var(\bar{y}).$$

$$var(\bar{y}) = N^2 \left(\frac{N-n}{N}\right) \cdot \frac{S^2}{n}.$$

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

Since the ratio estimate has the smaller variance, we are assuming, $Var(\hat{Y}_R) < var(\hat{Y})$.

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

Dividing RS_x; we have,

$$RS_x - 2\rho S_y < 0.$$

$$\frac{\bar{Y}}{\bar{X}} S_x - 2\rho S_y < 0$$

Dividing both sides by $\bar{Y}$.

$$\frac{S_x}{\bar{X}} < 2\rho \frac{S_y}{\bar{Y}} \Rightarrow \rho > \frac{1/2 \; S_x/\bar{X}}{S_y/\bar{Y}}.$$

$\Rightarrow \rho > 1/2$ co-efficient of variation of x H.T.P.

Theorem 4:

Bias in ratio estimate show that $\left|\frac{\text{Bias in } \hat{R}}{\hat{\sigma}_R}\right| \le$ C.V. of $\bar{x}$.

Proof:-

This result given an extract result for the bias and an upper bound to the ration of the bias to the standard error. consider the covariance in SRS of size of the quantitives $\hat{R}$ and $\bar{x}$ we have.

$$\text{co.v}(\hat{R}\bar{x}) = E(\hat{R}\bar{x}) - E(\hat{R})E(\bar{x})$$

$$= \frac{\bar{y}}{\bar{x}} \cdot \bar{x} - E(\hat{R})\bar{x}.$$

$$\therefore \text{co.v}(\hat{R}\bar{x}) = \bar{Y} - E(\hat{R})\bar{x}.$$

Now,

$$E(\hat{R}) = R - \frac{\text{cov}(\hat{R}\bar{x})}{\bar{x}}.$$

$$\text{Bias in } \hat{R} = \frac{-\text{co.v}(\hat{R}\bar{x})}{\bar{x}}$$

$$|\text{Bias in } \hat{R}| = \left|\frac{\text{co.v}(\hat{R}\bar{x})}{\bar{x}}\right| \quad \left[\because \rho = \frac{\text{co.v}(x,y)}{\sigma_x \sigma_y}. \quad \rho\sigma_x\sigma_y = \text{co.v}(x,y)\right]$$

$$= \left|\frac{\rho \hat{\sigma}_R \sigma_{\bar{x}}}{\bar{x}}\right|.$$

$$\therefore \frac{\text{Bias in } \hat{R}}{\sigma \hat{R}} = \frac{\sigma \bar{x}}{\bar{x}}.$$

$$= \text{co-variance}(\bar{x}).$$

$$\left|\frac{\text{Bias in } \hat{R}}{\sigma \hat{R}}\right| \le \text{covariance of } (\bar{x}).$$