Frequently Asked Questions

1. Prove that Variance of the estimated population mean under Stratified Random Sampling when Proportional allocation is

$$V(\bar{y}_{st})_{prop} = \sum_{h=1}^{k} \frac{W_h\ S_h^{\ 2}}{n} - \sum_{h=1}^{k} \frac{W_h\ S_h^{\ 2}}{N}$$

Answer:

We know that

$$V(\bar{y}_{st}) = \sum \frac{W_h^{\ 2} S_h^{\ 2}}{n_h} - \sum \frac{W_h^{\ 2} S_h^{\ 2}}{N_h}$$

Under P.A. we have

$$n_h = \frac{N_h}{N} * n \quad \text{and} \quad W_h = \frac{N_h}{N}$$

Then

$$V(\bar{y}_{st})_{p.a.} = \sum \frac{(\frac{N_h}{N})^2 S_h^{\ 2}}{\frac{N_h}{N} \cdot n} - \sum \frac{(\frac{N_h}{N})^2 S_h^{\ 2}}{N_h}$$

$$= \sum \frac{(\frac{N_h}{N}) S_h^{\ 2}}{n} - \sum \frac{(\frac{N_h}{N}) S_h^{\ 2}}{N}$$

$$= \sum \frac{W_h S_h^{\ 2}}{n} - \sum \frac{W_h S_h^{\ 2}}{N}$$

2. Obtain an expression for the Variance of the estimated population mean under Stratified Random Sampling when Optimum allocation

Answer:

We know that

$$V(\bar{y}_{st}) = \sum \frac{W_h^2 S_h^2}{n_h} - \sum \frac{W_h^2 S_h^2}{N_h}$$

Under O.A. we have

$$n_h = \frac{n}{\sum N_h S_h} \cdot N_h S_h \text{ and } W_h = \frac{N_h}{N}$$

Then

By substituting

$$V(\bar{y}_{st})_{o.a.} = \sum \frac{(\frac{N_h}{N})^2 S_h^2 \sum N_h S_h}{n N_h S_h} - \sum \frac{(\frac{N_h}{N})^2 S_h^2}{N_h}$$

$$= \sum \frac{N_h S_h \sum N_h S_h}{NNn} - \sum \frac{N_h S_h^2}{NN}$$

$$= \frac{(\sum W_h S_h)(\sum W_h S_h)}{n} - \frac{\sum W_h S_h^2}{N}$$

$$= \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^2}{N}$$

3.Show that Stratified sampling under proportional allocation is more efficient than SRSWOR

Answer:

We know that

$$S_h{}^2 = \frac{1}{N_h - 1} \cdot \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2$$

$$\Rightarrow (N_h - 1) S_h{}^2 = \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 \longrightarrow 1$$

Consider $S^2 = \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y})^2$

By substracting and adding $\overline{Y}_h$ withen summation we get,

$$S^2 = \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h + \overline{Y}_h - \overline{Y})^2$$

$$= \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (\overline{Y}_h - \overline{Y})^2 + \frac{2}{N-1} \sum_{h=1}^{k} \sum_{i=1}^{N_h} [(Y_{hi} - \overline{Y}_h)(\overline{Y}_h - \overline{Y})]$$

$$= \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (\overline{Y}_h - \overline{Y})^2 + \frac{2}{N-1} \sum_{h=1}^{k} (\overline{Y}_h - \overline{Y}) \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)]$$

$$= \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (\overline{Y}_h - \overline{Y})^2 \quad \left[\because \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h) = 0 \right]$$

From equation 1 we get

$$S^2 = \frac{1}{N-1} \cdot \sum_{h=1}^{k} (N_h - 1) S_h{}^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} N_h (\overline{Y}_h - \overline{Y})^2$$

When N-1 and N_h-1 are sufficiently large we can take $N - 1 \approx N$ and

$N_h - 1 \approx N_h$

$$S^2 = \frac{1}{N} \cdot \sum_{h=1}^{k} N_h S_h{}^2 + \frac{1}{N} \cdot \sum_{h=1}^{k} N_h (\overline{Y}_h - \overline{Y})^2$$

$$S^2 = \sum_{h=1}^{k} W_h S_h{}^2 + \sum_{h=1}^{k} W_h (\overline{Y}_h - \overline{Y})^2 \qquad \left[\because W_h = \frac{N_h}{N} \right]$$

Now by multiplying both LHS and RHS by $\frac{N-n}{Nn}$

$$[\frac{N-n}{Nn}]S^2 = \frac{N-n}{Nn}\sum_{h=1}^{k} W_h S_h^{\,2} + \frac{N-n}{Nn}\sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2$$

$$= \frac{\sum_{h=1}^{k} W_h S_h^{\,2}}{n} - \frac{\sum_{h=1}^{k} W_h S_h^{\,2}}{N} + \frac{N-n}{Nn}\sum_{h=1}^{k} W_h (\bar{Y}_h - \bar{Y})^2$$

But

$$v(\bar{y})_{wor} = [\frac{N-n}{Nn}]S^2$$

$$v(\bar{y})_{wor} = \frac{\sum_{h=1}^{k} W_h S_h^{\,2}}{n} - \frac{\sum_{h=1}^{k} W_h S_h^{\,2}}{N} +$$ some positive quantity.

$V(\bar{y})_{SRSWOR} = V(\bar{y})_{p.a.} +$ some positive quantity

$V(\bar{y})_{SRSWOR} \geq V(\bar{y})_{p.a.}$ OR

$V(\bar{y})_{p.a.} \leq V(\bar{y})_{SRSWOR}$

4. When the stratification is more useful?

Answer:

Stratification is most useful when the stratifying variables are

- simple to work with,
- easy to observe, and
- closely related to the topic of the survey.

5. Show that stratified Sampling using Optimum allocation is more efficient than Proportional Allocation.

Answer:

We can prove the above by proving

$$V(\bar{y})_{p.a.} - V(\bar{y})_{o.a.} \geq 0$$

$$V(\bar{y})_{p.a.} - V(\bar{y})_{o.a.} = \frac{\sum W_h S_h^{\,2}}{n} - \frac{\sum W_h S_h^{\,2}}{N} - \left[\frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^{\,2}}{N}\right]$$

$$= \frac{\sum W_h S_h^{\,2}}{n} - \frac{(\sum W_h S_h)^2}{n}$$

$$= \frac{\sum N_h S_h^{\,2}}{Nn} - \frac{(\sum N_h S_h)^2}{N^2 n}$$

$$= \frac{1}{Nn}\left[\sum N_h S_h^{\,2} - \frac{(\sum N_h S_h)^2}{N}\right]$$

$$= \frac{1}{Nn}\left[\sum N_h S_h^{\,2} - \frac{2(\sum N_h S_h)^2}{N} + \frac{(\sum N_h S_h)^2}{N}\right]$$

$$= \frac{\sum N_h}{Nn}\left[S_h^{\,2} - \frac{2S_h(\sum N_h S_h)}{N} + \frac{\sum N_h S_h^{\,2}}{N^2}\right]$$

$$= \frac{\sum N_h}{Nn}\left[S_h - \frac{\sum N_h S_h}{N}\right]^2$$

$$= \frac{\sum N_h}{Nn}[a-b]^2$$

Where $a = S_h \;\&\; b = \frac{\sum N_h S_h}{N}$ which is a positive quantity i.e.,

$$V(\bar{y})_{p.a.} - V(\bar{y})_{o.a.} \geq 0$$

$$\Rightarrow V(\bar{y})_{p.a.} \geq V(\bar{y})_{o.a.}$$

$$\Rightarrow V(\bar{y})_{o.a.} \leq V(\bar{y})_{p.a.}$$

6. Show that $V(\bar{y})_{o.a.} \leq V(\bar{y})_{p.a.} \leq V(\bar{y})_{SRSWOR}$

Answer:

We know that

$$V(\bar{y}_{st}) = \sum \frac{W_h^{\,2} S_h^{\,2}}{n_h} - \sum \frac{W_h^{\,2} S_h^{\,2}}{N_h}$$

Under O.A. we have

$$n_h = \frac{n}{\sum N_h S_h} \cdot N_h S_h \quad \text{and} \quad W_h = \frac{N_h}{N}$$

Then

By substituting

$$V(\bar{y}_{st})_{o.a.} = \sum \frac{(\frac{N_h}{N})^2 S_h^{\,2} \sum N_h S_h}{n N_h S_h} - \sum \frac{(\frac{N_h}{N})^2 S_h^{\,2}}{N_h}$$

$$= \sum \frac{N_h S_h \sum N_h S_h}{NNn} - \sum \frac{N_h S_h^{\,2}}{NN}$$

$$= \frac{(\sum W_h S_h)(\sum W_h S_h)}{n} - \frac{\sum W_h S_h^{\,2}}{N}$$

$$= \frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^{\,2}}{N}$$

Comparison of Optimum, Proportional and SRSWOR

P.T. $V(\bar{y})_{o.a.} \leq V(\bar{y})_{p.a.} \leq V(\bar{y})_{SRSWOR}$

Proof: We know that

$$S_h^{\,2} = \frac{1}{N_h - 1} \cdot \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2$$

$$\Rightarrow (N_h - 1)S_h^{\ 2} = \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 \longrightarrow 1$$

Consider $S^2 = \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y})^2$

By substracting and adding $\overline{Y}_h$ withen summation we get,

$$S^2 = \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h + \overline{Y}_h - \overline{Y})^2$$

$$= \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (\overline{Y}_h - \overline{Y})^2 + \frac{2}{N-1} \sum_{h=1}^{k} \sum_{i=1}^{N_h} [(Y_{hi} - \overline{Y}_h)(\overline{Y}_h - \overline{Y})]$$

$$= \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (\overline{Y}_h - \overline{Y})^2 + \frac{2}{N-1} \sum_{h=1}^{k} (\overline{Y}_h - \overline{Y}) \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)]$$

$$= \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h)^2 + \frac{1}{N-1} \cdot \sum_{h=1}^{k} \sum_{i=1}^{N_h} (\overline{Y}_h - \overline{Y})^2 \quad \left[\because \sum_{i=1}^{N_h} (Y_{hi} - \overline{Y}_h) = 0 \right]$$

From equation 1 we get

$$S^2 = \frac{1}{N-1} \cdot \sum_{h=1}^{k} (N_h - 1)S_h^{\ 2} + \frac{1}{N-1} \cdot \sum_{h=1}^{k} N_h (\overline{Y}_h - \overline{Y})^2$$

When N-1 and N_h-1 are sufficiently large we can take $N - 1 \approx N$ and $N_h - 1 \approx N_h$

$$S^2 = \frac{1}{N} \cdot \sum_{h=1}^{k} N_h S_h^{\ 2} + \frac{1}{N} \cdot \sum_{h=1}^{k} N_h (\overline{Y}_h - \overline{Y})^2$$

$$S^2 = \sum_{h=1}^{k} W_h S_h^{\ 2} + \sum_{h=1}^{k} W_h (\overline{Y}_h - \overline{Y})^2 \qquad \left[\because W_h = \frac{N_h}{N} \right]$$

Now by multiplying both LHS and RHS by $\frac{N-n}{Nn}$

$$[\frac{N-n}{Nn}]S^2 = \frac{N-n}{Nn} \sum_{h=1}^{k} W_h S_h^{\ 2} + \frac{N-n}{Nn} \sum_{h=1}^{k} W_h (\overline{Y}_h - \overline{Y})^2$$

$$= \frac{\sum_{h=1}^{k} W_h S_h^{\;2}}{n} - \frac{\sum_{h=1}^{k} W_h S_h^{\;2}}{N} + \frac{N-n}{Nn} \sum_{h=1}^{k} W_h (\overline{Y}_h - \overline{Y})^2$$

But

$$v(\overline{y})_{wor} = [\frac{N-n}{Nn}]S^2$$

$$v(\overline{y})_{wor} = \frac{\sum_{h=1}^{k} W_h S_h^{\;2}}{n} - \frac{\sum_{h=1}^{k} W_h S_h^{\;2}}{N} + \text{some positive quantity.}$$

$V(\bar{y})_{SRSWOR} = V(\bar{y})_{p.a.} +$ some positive quantity (from theorem 1)

$V(\bar{y})_{SRSWOR} \geq V(\bar{y})_{p.a.}$ OR

$V(\bar{y})_{p.a.} \leq V(\bar{y})_{SRSWOR}$ $\longrightarrow$ 2

Now let us prove that $V(\bar{y})_{o.a.} \leq V(\bar{y})_{p.a.}$

We can prove the above by proving

$$V(\bar{y})_{p.a.} - V(\bar{y})_{o.a.} \geq 0$$

$$V(\bar{y})_{p.a.} - V(\bar{y})_{o.a.} = \frac{\sum W_h S_h^{\;2}}{n} - \frac{\sum W_h S_h^{\;2}}{N} - \left[\frac{(\sum W_h S_h)^2}{n} - \frac{\sum W_h S_h^{\;2}}{N} \right]$$

$$= \frac{\sum W_h S_h^{\;2}}{n} - \frac{(\sum W_h S_h)^2}{n}$$

$$= \frac{\sum N_h S_h^{\;2}}{Nn} - \frac{(\sum N_h S_h)^2}{N^2 n}$$

$$= \frac{1}{Nn} \left[\sum N_h S_h^{\;2} - \frac{(\sum N_h S_h)^2}{N} \right]$$

$$= \frac{1}{Nn}\left[\sum N_h S_h^{\ 2} - \frac{2(\sum N_h S_h)^2}{N} + \frac{(\sum N_h S_h)^2}{N}\right]$$

$$= \frac{\sum N_h}{Nn}\left[S_h^{\ 2} - \frac{2S_h(\sum N_h S_h)}{N} + \frac{\sum N_h S_h^{\ 2}}{N^2}\right]$$

$$= \frac{\sum N_h}{Nn}\left[S_h - \frac{\sum N_h S_h}{N}\right]^2$$

$$= \frac{\sum N_h}{Nn}[a-b]^2$$

Where $a = S_h \;\&\; b = \frac{\sum N_h S_h}{N}$ which is a positive quantity i.e.,

$$V(\bar{y})_{p.a.} - V(\bar{y})_{o.a.} \geq 0$$

$$\Rightarrow V(\bar{y})_{p.a.} \geq V(\bar{y})_{o.a.}$$

$$\Rightarrow V(\bar{y})_{o.a.} \leq V(\bar{y})_{p.a.} \longrightarrow 3$$

By comparing equations 2 and 3 we get

$$V(\bar{y})_{o.a.} \leq V(\bar{y})_{p.a.} \leq V(\bar{y})_{SRSWOR}$$

7. Why Stratified sampling is a suitable sampling technique?

Answer:
Stratified sampling is suitable in most of the situations because of the following reasons

- Gain in precision – Stratified sampling gives estimates of higher precision
- Flexible in the choice of the sample design for different strata- We can use any sampling design for different strata
- Able to get estimates of each stratum in addition to the population estimates

8. When the sample survey leads to good survey results?

Answer:

Always when the sample is properly designed leads to good survey results

• When conducting the survey and the target population shows subpopulations, it is highly recommended to use Stratified Random Sampling to design sample compared to SRS

• Stratification will give us better estimates for the population parameters compared to those we would get using Simple Random Sampling

9. How do you measure the efficiency of any sample design as compared to simple random sample?

Answer:

Measures used to estimate how efficient is the sample design compared to the Simple Random Sample are

D= variance of any estimator for a given design/ variance of any estimator

for SRS

• If D < 1, then the given Design is more efficient than the SRS.

• If D> 1, then the given Design is less efficient than the SRS.

10. Compare unstratified sampling with that of stratified sampling.

Answer:

Compared to unstratified sampling, stratified sampling

(1) permits the estimation of population parameters and within-strata inferences and comparisons across strata;

(2) tends to be more representative of a population;

(3) takes advantage of knowledge the researcher has about the population;
(4) possibly makes for lower data collection costs; and

(5) permits the researcher to use different sampling procedures within the strata

11. How the estimates obtained using stratified random sampling is of more precision than that of Simple Random sampling?

Answer:

In stratified random sampling since we divide the entire heterogeneous population into layers of homogeneous characteristics variations within the strata is minimized.As a consequence of the reduction in the variability within each stratum stratified random sampling provides more efficient estimates as compared with simple random sampling. For example: the sample estimate of the population mean is more efficient in both proportional and Neyman's allocation of the samples to different strata in stratified random sampling as compared with the corresponding estimates obtained in SRS

Sometimes it is desired to achieve different degrees of accuracy for different segments of the population. Stratified random sampling is the only sampling plan which enables us to obtain the results of known precision for each of the stratum.

12. How do you measure the efficiency of Proportional Allocation and Optimum Allocation with that of SRSWOR?

Answer:

Efficiency of 3 techniques:

Efficiency of SRSWOR w.r.t. P.A. is given by

$$\frac{V(\bar{y})_{SRSWOR}}{V(\bar{y})_{p.a.}} \times 100$$

Efficiency of O.A. w.r.t. SRSWOR is given by

$$\frac{V(\bar{y})_{SRSWOR}}{V(\bar{y})_{o.a.}} \times 100$$

13. What are the limitations of Stratified Random sampling?

Answer:

1) As already pointed out the success of the stratified sampling depends on
   i) Effective stratification of the universe into homogeneous strata
   ii) Appropriate size of the samples to be drawn from each stratum

   If Stratification if faulty the results will be biased. The error due to wrong stratification cannot be compensated even by taking large samples

   The allocation of the sample sizes to different strata requires an accurate knowledge of the population size in each stratum N_h , h = 1,2,..,k. Neyman's principle of Optimum allocation n_h proportional to N_hS_h requires addition knowledge of the variability or S.D of each strata . N_h and S_h are usually unknown and are the serious limitations of effective use of stratified random sampling
2) It is a very difficult task to divide the universe into homogeneous strata
3) If the strata are overlapping , unsuitable or disproportionate the selection of the samples may be biased. Such errors cannot be compensated even by taking large samples.
4) Disproportionate stratification requires weighting which again introduces selective factor in the sample and under weighing makes the sample unrepresentative.

   On the other hand, unlike unstratified sampling, stratified sampling requires prior information on the stratification variables and more complex analysis procedures.

14. Give an expression for gain in efficiency due to P.A and O.A with that of SRSWOR.

Answer:

Gain in efficiency of P.A. due to SRSWOR is given by

$$\left[\frac{V(\bar{y})_{SRSWOR} - V(\bar{y})_{p.a.}}{V(\bar{y})_{p.a.}}\right] \times 100$$

Gain in efficiency of O.A. due to SRSWOR is given by

$$\left[\frac{V(\bar{y})_{SRSWOR} - V(\bar{y})_{o.a.}}{V(\bar{y})_{o.a.}}\right] \times 100$$

15. What are the conditions required to be satisfied for Stratified sampling to acquire large gain in precision?

Answer:

Stratification will produce large gains in precision by satisfying the following three conditions

1) The population is composed of subgroups varying widely in size
2) The principal variables to be measured are closely related to the sizes of the subgroups
3) A good measure of size is available for setting up the strata