Frequently Asked Questions

1. Write a note on Systematic Sampling.

Answer:
Systematic sampling is a technique which has a nice feature of selecting a whole sample with just one random start. A sampling technique in which first unit is selected with a help of random numbers and the others get selected automatically according to some pre-designed pattern is known as systematic random sampling. This is a commonly employed technique if the complete and up-to -date list of the sampling units is available.

2. How do you select sample observations under Systematic sampling?

Answer:
We suppose that the population is linear in order in some way such that units can be referred to by number. Further let N be expressible in the form of **N=nk** and let the selected random number be **r (<k), k** being called as a sampling interval. In this procedure a sample comprise of the unit **r, r+k, r+2k, …, r+(n-1)k . T**he technique will generate **k** systematic samples with equal probability which may be shown as follows:

Hence Systematic sampling is a procedure of selecting every k^{th} unit starting from the unit corresponding to a number r selected at random from 1 to k, k being an integer nearest to N/n i.e. k=N/n => N=nk.

3. List down the basic concepts required for the estimation of parameters under systematic sampling.

Answer:

1) The number r is known as the random start. If r is a random start then a systematic sample of size n consists of the units r, r+k, r+2k, r+3k, …, r+(n-1)k.

2) Under systematic sampling we can have k possible samples and probability of selecting a sample is 1/k.

3) Under systematic sampling let y_{ri} be the i^{th} unit corresponding to the r^{th} sample r=1,2,....k , i=1,2,....,n

4. What are the notations used in Systematic Sampling?

Answer:

Mean of the rth systematic sample $\bar{y}_r = \dfrac{\sum_{i=1}^{n} y_{ri}}{n} \Rightarrow \sum y_{ri} = n\bar{y}_r$

Under systematic sampling

The Population Mean

$$\overline{Y} = \frac{1}{nk}\sum_{r=1}^{k}\sum_{i=1}^{n} Y_{ri} = \frac{1}{k}\sum_{r=1}^{k}\overline{y_r}$$ [covers whole population]

Population variance

$$\sigma^2 = \frac{1}{nk}\sum_{r=1}^{k}\sum_{i=1}^{n}(Y_{ri} - \overline{Y})^2 = \frac{1}{N}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \overline{Y})^2$$

Population Mean square

$$S^2 = \frac{1}{nk-1}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \overline{Y})^2 = \frac{1}{N-1}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \overline{Y})^2$$

5. Under Systematic sampling prove that $\bar{y}_r$ is an unbiased estimator of the population mean $\bar{Y}$

Answer:

Consider $E(\bar{y}_r) = \sum_r \bar{y}_r . p(\bar{y}_r)$

$$= \sum_r \bar{y}_r . \frac{1}{k}$$

$$= \sum_r \left[\frac{1}{n}\sum y_{ri}\right] . \frac{1}{k}$$

$$= \frac{1}{nk}\left[\sum_r \sum_i y_{ri}\right].$$

$$= \bar{Y}$$

$$\therefore E(\bar{y}_r) = \bar{Y}$$

Thus, if N=nk, the systematic sample mean provides an unbiased estimate of the population mean.

6. Under Systematic sampling prove that:

$$V(\bar{y}_r) = (\frac{N-1}{N})S^2 - (\frac{n-1}{n})S^2{}_{sym}$$

$$S^2{}_{sym} = \frac{1}{k(n-1)}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r)^2$$

Where

Answer:

$$S^2{}_{sym} = \frac{1}{k(n-1)}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r)^2$$

is the mean square among units which lie within the same systematic sample. The denominator of this variance , k(n-1) is constructed by the usual rules in the analysis of variance each of the k samples contributes (n-1) degrees of freedom to the sum of squares in the numerator.

$$S^2 = \frac{1}{nk-1}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y})^2$$

$$= \frac{1}{nk-1}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r + \bar{Y}_r - \bar{Y})^2$$

$$= \frac{1}{nk-1}\sum_{r=1}^{k}\sum_{i=1}^{n}\left[(y_{ri} - \bar{Y}_r)^2 + (\bar{Y}_r - \bar{Y})^2 + 2(y_{ri} - \bar{Y}_r)(\bar{Y}_r - \bar{Y})\right]$$

$$= \frac{1}{nk-1}\left[\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r)^2 + \sum_{r=1}^{k}\sum_{i=1}^{n}(\bar{Y}_r - \bar{Y})^2 + 2\sum_{r=1}^{k}(\bar{Y}_r - \bar{Y})\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r)\right]$$

$$= \frac{1}{nk-1}\left[\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r)^2 + \sum_{r=1}^{k}\sum_{i=1}^{n}(\bar{Y}_r - \bar{Y})^2\right]$$

$$(nk-1)S^2 = \left[\sum_{r=1}^{k} \sum_{i=1}^{n} (y_{ri} - \overline{Y}_r)^2 + \sum_{r=1}^{k} \sum_{i=1}^{n} (\overline{Y}_r - \overline{Y})^2 \right]$$

$$(nk-1)S^2 = \left[k(n-1)S^2{}_{sym} + \sum_{r=1}^{k} n(\overline{Y}_r - \overline{Y})^2 \right] \longrightarrow 1$$

We know that

$$V(\overline{y}_r) = E(\overline{y}_r - E(\overline{y}_r))^2$$

$$= E(\overline{y}_r - \overline{Y})^2$$

$$= \sum_r (\overline{y}_r - \overline{Y})^2 . \frac{1}{k}$$

$$k.V(\overline{y}_r) = \sum_r (\overline{y}_r - \overline{Y})^2 \longrightarrow 2$$

By substituting equation 2 in equation 1

$$(nk-1)S^2 = \left[k(n-1)S^2{}_{sym} + nkV(\overline{y}_r) \right]$$

$$\Rightarrow \frac{(nk-1)S^2 - k(n-1)S^2{}_{sym}}{nk} = V(\overline{y}_r)$$

$$\Rightarrow \frac{(nk-1)S^2}{nk} - \frac{k(n-1)S^2{}_{sym}}{nk} = V(\overline{y}_r)$$

$$\Rightarrow \frac{(N-1)S^2}{N} - \frac{(n-1)S^2{}_{sym}}{n} = V(\overline{y}_r)$$

$$\Rightarrow V(\overline{y}_r) = \frac{(N-1)S^2}{N} - \frac{(n-1)S^2{}_{sym}}{n}$$

7. What do you mean by intra class correlation coefficient in Systematic Sampling?

Answer:
Intra-class correlation coefficient is the correlation between the units of the same systematic sample and is denoted by ρ

The intra-class coefficient ρ

$$\rho = \frac{E(Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{E(Y_{ri} - \overline{Y})^2}$$

$$= \frac{\sum_{r=1}^{k} \sum_{i \neq j}^{n} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y}) . \frac{1}{kn(n-1)}}{\sum_{r=1}^{k} \sum_{i=1}^{n} (Y_{ri} - \overline{Y})^2 \frac{1}{kn}}$$

$$= \frac{\sum_{r} \sum_{i \neq j} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{\sum_{r} \sum_{i} (Y_{ri} - \overline{Y})^2 .(n-1)}$$

$$\rho = \frac{\sum_{r} \sum_{i \neq j} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{nk\sigma^2 (n-1)} \qquad \left[\boxtimes\ \sigma^2 = \frac{\sum_{r} \sum_{i \neq j} (Y_{ri} - \overline{Y})^2}{nk} \right]$$

8. Prove that Variance of sample mean under systematic sampling in terms of intra class correlation coefficient is $V(\bar{y}_r) = \frac{\sigma^2}{n}[1 + (n-1)\rho]$

   Where ρ is an intra-class correlation coefficient between the units of the same systematic sample?

Answer:

The intra-class coefficient ρ

$$\rho = \frac{E(Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y})}{E(Y_{ri} - \overline{Y})^2}$$

$$= \frac{\sum_{r=1}^{k} \sum_{i \neq j}^{n} (Y_{ri} - \overline{Y})(Y_{rj} - \overline{Y}) . \frac{1}{kn(n-1)}}{\sum_{r=1}^{k} \sum_{i=1}^{n} (Y_{ri} - \overline{Y})^2 \frac{1}{kn}}$$

$$= \frac{\sum_{r}\sum_{i\neq j}(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y})}{\sum_{r}\sum_{i}(Y_{ri}-\overline{Y})^2.(n-1)}$$

$$\rho = \frac{\sum_{r}\sum_{i\neq j}(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y})}{nk\sigma^2(n-1)} \qquad \left[⊠\ \sigma^2 = \frac{\sum_{r}\sum_{i\neq j}(Y_{ri}-\overline{Y})^2}{nk} \right]$$

$$\Rightarrow \sum_{r=1}^{k}\sum_{i\neq j}^{n}(Y_{ri}-\overline{Y})(Y_{rj}-\overline{Y}) = \rho nk\sigma^2(n-1) \longrightarrow 1$$

Consider $V(\overline{y}_r) = E(\overline{y}_r - E(\overline{y}_r))^2$

$$= E(\overline{y}_r - \overline{Y})^2$$

$$= \sum_{r=1}^{k}(\overline{y}_r - \overline{Y})^2 \cdot \frac{1}{k}$$

$$= \frac{1}{k}\sum_{r=1}^{k}(\frac{1}{n}\sum_{i} y_{ri} - \overline{Y})^2$$

$$= \frac{1}{k}\sum_{r}(\frac{1}{n}\sum_{i=1}^{n} y_{ri} - \overline{Y})^2.$$

$$= \frac{1}{n^2k}\sum_{r}(\sum_{i=1}^{n} y_{ri} - \overline{Y})^2.$$

$$= \frac{1}{n^2k}\sum_{r=1}^{k}\left[\sum_{i=1}^{n}(y_{ri}-\overline{Y})^2 + \sum\sum_{i\neq j}(y_{ri}-\overline{Y})(y_{rj}-\overline{Y})\right]$$

$$= \frac{1}{n^2k}[\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri}-\overline{Y})^2 + \sum_{r=1}^{k}\sum_{i\neq j}^{n}(y_{ri}-\overline{Y})(y_{rj}-\overline{Y})]$$

$$V(\overline{y}_r) = \frac{1}{n^2k}[nk\sigma^2 + nk(n-1)\rho\sigma^2]$$ [from equation 1]

$$V(\bar{y}_r) = \frac{\sigma^2}{n}[1+(n-1)\rho]$$

9. How do you interpret the variability of the estimate in terms of intra class correlation?

Answer:

Thus, we see that a positive intraclass correlation between the units of the same sample inflates the variability of the estimate. Even a small positive correlation may have a larger effect. Due to the multiplier

(n-1) , the increase is quite significant even for small values of ρ

We have $V(\bar{y}_r) \geq 0 \Rightarrow \rho \geq \frac{-1}{n-1}$

Thus the minimum value of ρ is $\frac{-1}{n-1}$ and in this case $V(\bar{y}_r) = 0$

10. How can you express the variance of the systematic sample interms of S2 and intra class correlation coefficient?

Answer:

Variance of sample mean under systematic sampling in terms of intra class correlation coefficient can also be obtained as

$$V(\bar{y}_r) = \frac{(nk-1)S^2}{n^2k}[1+(n-1)\rho]$$

Where ρ is an intra-class correlation coefficient between the units of the same systematic sample?

In this case Rho can be expressed as:

$$\rho = \frac{\sum_{r}\sum_{i\neq j}(Y_{ri}-\bar{Y})(Y_{rj}-\bar{Y})}{(nk-1)S^2(n-1)} \qquad \left[\boxtimes S^2 = \frac{\sum_{r}\sum_{i\neq j}(Y_{ri}-\bar{Y})^2}{nk-1} \right]$$

11. How can you relate the variance of the systematic sample to that of SRS?

Answer:

Since variance of a systematic sample mean is expressed in terms of S^2 we can relate it to the variance for a simple random sample Systematic sample is precise when units within the same sample are heterogeneous and imprecise when units within the same sample are homogeneous. If there is little variation within a systematic sample relative to that in the population the successive units in the sample are repeating more or less the same information.

12. What do you mean by Standard error?

Answer:

Standard error can be defined as the Standard deviation of the distribution

For example, suppose t is an estimator then Standard Error of t is equal to square root of variance of t i.e., the Standard deviation of any estimator gives the Standard error of that estimator.

13. What are the purposes of obtaining standard errors of an estimator?

Answer:

The formulae for the standard errors of the estimated population mean and total are used primarily for three purposes:

1. To compare the precision obtained by Systematic sampling with that given by other methods of sampling
2. To estimate the size of the sample needed in a survey that is being planned
3. To estimate the precision actually attained in a survey that has been completed.

14. What is the Standard Error of an estimated population mean under Systematic sampling?

Answer:

Standard error of estimated population mean

$$\text{S.E}(\bar{y}_r) = \sqrt{V(\bar{y}_r)} = \sqrt{\frac{N-1}{N}S^2 - \frac{n-1}{n}S^2_{sys}}$$

$$S^2{}_{sym} = \frac{1}{k(n-1)}\sum_{r=1}^{k}\sum_{i=1}^{n}(y_{ri} - \bar{Y}_r)^2$$

where

15. Give an expression for the Standard error of the systematic sample mean in terms of intra -class correlation.

Answer:

Standard Error if the estimated population mean can also be expressed in terms of intra-class correlation

$$S.E\,(\bar{y}_r) = \sqrt{V(\bar{y}_r)} = \sqrt{\frac{\sigma^2}{n}[1 + (n-1)\rho]}$$

OR in terms of S^2

$$S.E\,(\bar{y}_r) = \sqrt{V(\bar{y}_r)} = \sqrt{\frac{(nk-1)S^2}{n^2k}[1 + (n-1)\rho]}$$