Data collection, Analysis and Inference

Subject Code: CPE-RPE,

May 2021,
SRM Univeristy-AP, Andhrapradesh

Lecture- 6: Sample size determination & sampling techniques: Systematic sampling, Cluster sampling

- Aim: To implement the methods of Systematic sampling and through forming clusters.

Systematic sampling

Definition: Systematic sampling is defined as a probability sampling method where the researcher chooses elements from a target population by selecting a random starting point and selects sample members after a fixed ‘sampling interval.’

The Technique

Step 1: Assign a number to every element in your population (indexing).

Step 2: Decide how large your sample size should be.

Step 3: Define the interval of this sample. This will be the standard distance between the elements.

Step 4: Divide the population by the sample size.

Step 5: Randomly choose the starting member (r) of the sample and add the interval to the random number to keep adding members in the sample.

Example:

- For a simple example, let's say you have a population of 100 people, so you'll assign the numbers 1 to 100 to the group.

  Let's say you need a sample of 10 people.

  So, the interval=100 / 10 = 10.

  This is your "nth" sampling digit (i.e. you'll choose every 10^{th} item)

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40
41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80
81	82	83	84	85	86	87	88	89	90
91	92	93	94	95	96	97	98	99	100

- If the starting point is 10 (Randomly picked) then the sample is S={10,20,30,40,50,60,70,80,90,100}.
- If the starting point is 4 (Randomly picked) then the sample is S={4,14,24,34,44,54,64,74,84,94}.

Cluster sampling

It is one of the basic assumptions in any sampling procedure that the population can be divided into a finite number of distinct and identifiable units, called **sampling units**.

- The smallest units into which the population can be divided are called elements of the population. The groups of such elements are called clusters.

Construction of clusters

The clusters are constructed such that the sampling units are heterogeneous within the clusters and homogeneous among the clusters.

Note: This is opposite to the construction of the strata in the stratified sampling.

How to do it ?

In cluster sampling

- divide the whole population into clusters according to some well-defined rule.

- Treat the clusters as sampling units.

- Choose a sample of clusters according to some procedure (Simple or Systematic).

- Carry out a complete enumeration of the selected clusters, i.e., collect information on all the sampling units available in selected clusters.

Case of equal clusters (to estimate population mean)

- Suppose the population is divided into N clusters and each cluster is of size M .
- Select a sample of n clusters from N clusters by the method of SRS, generally WOR.

  So
- total population size = NM
- total sample size = nM .

- Let y_{ij} : Value of the characteristic under study for the value of j^{th} element (j= 1, 2,..., M) in the i^{th} cluster (i= 1, 2,..., N).

$$\bar{y}_i = \frac{1}{M}\sum_{j=1}^{M} y_{ij} \quad \text{mean per element of } i^{th} \text{ cluster .}$$

Consider the mean of all such cluster means as an estimator of **population mean** as

$$\bar{y}_{cl} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i$$

Case of unequal clusters:

- Let there be N clusters and M_i be the size of i^{th} cluster, let

$$M_0 = \sum_{i=1}^{N} M_i$$

$$\bar{M} = \frac{1}{N}\sum_{i=1}^{N} M_i$$

$$\bar{y}_i = \frac{1}{M_i}\sum_{j=1}^{M_i} y_{ij} \text{ : mean of } i^{th} \text{ cluster}$$

$$\bar{Y} = \frac{1}{M_0}\sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij}$$

$$= \sum_{i=1}^{N} \frac{M_i}{M_0}\bar{y}_i$$

$$= \frac{1}{N}\sum_{i=1}^{N} \frac{M_i}{\bar{M}}\bar{y}_i$$

- END