Chapter 3

Multistage Sampling: Two-Stage Equal Cluster Sampling

Multi-stage Sample Design

- For studies of **large and geographically dispersed** populations it is more convenient to use a **multi-stage sampling design.**
- It is particularly appropriate
  - where a **large scale survey** is to be conducted, and
  - where for **logistic and organizational reasons** it is convenient for the sample to be grouped together in a **more limited number of geographical areas**, rather than being spread thinly and **dispersed across** the whole country.

- Multistage sampling is adopted in a number of situations:
  - Sampling frames may not be available for all the ultimate observational units in the entire population.
  - A multistage sampling plan may be more convenient than a single stage sample of the ultimate units, as the cost of surveying and supervising;
    - in large scale survey can be very high due to travel, identification and contact, etc.
  - It can be convenient means of reducing response errors and improving sampling efficiency by reducing the ***intra-class correlation coefficient*** observed in natural sampling units.

- In un-stratified multistage sampling, the sample is selected in stages,
  - i.e., the population is divided in to a number of (primary sample units)PSUs, which are sampled; then
  - the selected PSUs are sub divided into a number of smaller second stage units, which are again sampled;
  - the process is continued until the ultimate sampling units are reached.

- For multistage simple random sampling, at each stage the selection design is SRS, with an equal selection probability for each stage.
  - For example, for two stage simple random sample the selection method is SRS at first stage with equal probability(1/total PSUs) and
  - SRS at second – stage again with equal probability(1/ total SSUs)in which the method is described in short as SRS/SRS.

- For a multistage varying probability sampling with two stage designs, the selection method could be
  - probability proportional to size either at both sampling stages (PPS/PPS) or
  - PPS at first stage and SRS at second stage (PPS/SRS) and
  - similar procedure can be followed for more than two stages.
- A basic principle of scientific sampling is that every sampling unit must have a known, positive probability of being selected.
- Where the probabilities are equal, the sample design is known **self-weighting** and
- the formulae for calculating estimates are relatively straight forward.
- Where the sample design is **not self-weighting**, then the data relating to different sample units have to be weighted.
- At both stages simple random sample is used (SRS).

Simple random sampling (SRS):

- It is the simplest kind of sampling method.
- It requires as a sampling frame a list of sampling units- households, farmers, institutions, or whatever else is being used – in any convenient order.
- The items listed must be numbered in sequence, starting from one the first item at the head of the list and continuing up to as many as the there are items listed.

- A table of random numbers is needed to obtain a random selection of these numbers, and the items, which have been given the selected numbers that form the sample chosen for the survey.
  - The use of random numbers ensures that the sample units are chosen entirely by chance, without being influenced by any person's unconscious preferences.
  - In a table of random numbers, each number within the range has an equal chance or probability of selection.
  - Since each element in the sample frame is given one number, each unit has an equal chance of selection for the sample.
  - The sampling could be performed with or without replacement.

Linear systematic sample (LSS):

- it is operationally a convenient method of selecting a sample.
- In a systematic sample we decide the sample size n from a population of size N. In this case, however, the population has to be organized in some way,
  - such as points along a path or in simple numerical order.
  - We choose a starting point the sequence by selecting the rth unit from one' end 'of the sequence, where r is less than n. and is usually chosen between 1 and k randomly.
  - We then take the rest of the sample by adding K to r, where k is an integer number equal to N/n or to the next lowest integer below N/n if this division produces a real number. We do this repeatedly until we reach the end of sequence. One way of envisioning a systematic sample is think of the sample frame as a 'row; of units, and the sample as a sequence of equal of equal-spaced 'stops' along the row, as shown below-
- Let $N = population size,\ n = sample\ size,\ k = \frac{N}{n} = sampling\ interval.$

Chose any number between 1 and K. let rth

- The rth unit is selected at first and then rth+k until the required sample size is reached.

Cluster sampling:

- clusters can be defined as sampling units containing several elements that occur in groups naturally or formed artificially.
- A cluster has listing units associated with it in which the units can be
  - geographical,
  - temporal (chronological,in time etc.) or
  - spatial in nature.
- Thus, cluster sampling can be defined as any sampling plan that uses a frame consisting of clusters of listing units.
- Clusters are formed in a way that elements with in a cluster are heterogeneous,
  - i.e. observations in each cluster should be more or less dissimilar.
- In a single stage sampling, we select a sample of clusters and completely cover all units within selected clusters.
- Clusters can be selected by a variety of sampling techniques.
  - For example, we can select a sample of clusters by SRS or by systematic sampling or sampling with PPS.

- The important reasons for using cluster sampling are feasibility and economy.
- If the only sampling frames readily available for the target population are lists of clusters, then the only feasible method of sampling is cluster sampling.
  - That is why for surveys of human population; to compile lists of households for the purpose of survey never seems feasible in terms of time and resources.
  - Listing costs and traveling costs almost always lowest for cluster sampling.
- The disadvantage of cluster sampling is that the **standard errors** of estimates obtained form this design are often **high** compared with those obtained from samples of the same number of listing units chosen by other sampling designs.
  - Therefore, one can choose the sampling design that gives the lowest possible standard errors at a specified cost or, conversely, the sampling design that yield, at the lowest cost, estimates having standard errors.

Stratified random sampling:

- Definition:

  The process of splitting the sample to take account of possible sub-populations is called stratification, and such techniques are called stratified sampling methods.

- Stratified sampling is a technique, which uses any relevant information, which might be available, in order to increase efficiency.
- Stratified sampling involves the division or stratification of a population by partitioning the sampling frame into non-overlapping and relatively homogeneous groups called strata.
- The selection of samples can be performed independently in each of those strata.

Stratified random sampling...

- Stratified random sampling is a sampling plain in which a population is dividing into L mutually exclusive and exhaustive strata, and
- a simple random sample of nh elements is taken separately and independently within each stratum.
- Let $N_1, N_2, ..., N_L$ represent the number of sampling units within each stratum, and n1,n2,..nL represent the number of randomly selected sampling units within each stratum. Then the total number of possible stratified random samples is equal to

$$\binom{N_1}{n_1} \times \binom{N_2}{n_2} \times ... \times \binom{N_L}{n_L} \leq \binom{N}{n}$$

Stratified random sampling...

- Stratified random sampling, in particular, involves dividing the population into strata, and
- then selecting simple random samples from each of the strata.
- Stratification variables may be
  - geographic (region, province, rural urban, zone) or
  - non-geographic (income, age, sex, size of employees, etc).
- It should be kept in mind that stratification is limited only to those items of information, which are available on the frame.

The Purpose of Stratified Sampling

- Stratified sampling is used in certain types of surveys because
  - it combines the conceptual simplicity of simple random sampling with potentially significant gains in reliability.
- Basically there are four major reasons for resorting to stratification:

1. The principal objective of stratification is to reduce sampling error.
   - Under certain conditions the variances of the sample estimates may be decreased, which means precision may be increased over simple random sample.
   - In a stratified sample, the sampling error depends on the population variance existing within the strata but not between the strata.
   - For this reason it prays to create strata with low internal variability.
   - It follows that the more homogeneous the groups, the greater the precision of the sample estimate.

2. In some cases, separate estimates are required at the stratum level. For example,in household surveys estimates may be required by
   - ❖ province,
   - ❖ income group,
   - ❖ occupation,
   - ❖ age group,
   - ❖ urban size group,
   - ❖ educational category, etc.
3. Stratified sampling is administratively convenient.
   - ❖ It can enable a survey organization to control the distribution of fieldwork among its regional offices.
   - ❖ Also, for large complex surveys, it can facilitate sample design work by enabling such work to be carried out within operationally manageable units.

4. Sometimes, different parts of the population may call for different sampling procedures.

- For example :
  - here may be incomplete frames in which lists may be available for different parts of the population and in this case supplementing an incomplete list will be required.
  - Physical distribution of a population such as private households, institutions (military camp, prisons and hostel), densely or sparsely populated areas may call for different procedures.
  - A different procedure may be used to sample persons in sparsely populated rural areas than that used for the more densely populated urban areas.
- The diverse nature of the sampling units of the population may call for different procedures.
  - For instance, employee's study requires stratification by literate and illiterate; and study on young people between 14 and 18 years requires young people at school and those who are not at school.

- The major disadvantage of stratified sampling is that it may take more time to select the sample than would be the case for simple random sampling.
- More time is involved because complete frames are necessary within each of the strata and each stratum must be sampled.