Sampling

Sampling: Sampling is the selection of a few units from all the observational units in a population, and a sample is a portion of the population selected to stud the characteristics of the whole population. For example, if the are 100 fishes in an aquarium and we decide to take the weight of 25 of them, then the weights of these 25 fishes constitute the sample while the weights of all 100 fishes is the population. Thus, sample should be a true representative of the whole population.

It is not possible to include each member of the population in an experimental study to find the prevalence of tuberculosis or to test the efficacy of a drug in all the patients suffering from a particular disease. Collecting data from individuals involves a large number of investigators hung amount of money, long time and much effort. Because of all such difficulties, we prefer to use an appropriate sampling technique.

In medical studies, the sampling data are collected from a population sufficiently large and representative of the population under study, chosen by a standard sampling technique. The word sample means a portion or a part of the population selected in some manner.

Types of sampling methods: There are two types of sampling techniques:

- **Random or probability sampling**
  - Simple random sampling
  - Systematic random sampling
  - Stratified random sampling
  - Multistage sampling
  - Multiphase sampling
  - Cluster sampling
- **Nonrandom or nonprobability sampling**
  - Convenience sampling
  - Purposive sampling
  - Quota sampling

Random Sampling: In random sampling a sample is selected in such a way that each and every element in the population has an equal chance of being included in the sample. A random sample must be selected in such a way that every element in the population has an equal opportunity of being included in the sample.

Simple Random Sampling: In this method samples are chosen at random and each member or sample unit of the population has an equal chance of being selected in the sample. This method is well applicable when the population is small, homogeneous and readily available, e.g. fishes present in the aquarium. In this sampling method every unit of the population has an

equal chance of being selected. or we can say that in other words, Simple random sampling (SRS) is a method of selection of a sample comprising of n number of sampling units out of the population having N number of sampling units such that every sampling unit has an equal chance of being chosen.

The samples can be drawn in two possible ways.

(i): The sampling units are chosen without replacement in the sense that the units once chosen are not placed back in the population .

(ii): The sampling units are chosen with replacement in the sense that the chosen units are placed back in the population.

Simple random sampling without replacement (SRSWOR): SRSWOR is a method of selection of n units out of the N units one by one such that at any stage of selection, anyone of the remaining units have same chance of being selected, i.e. $1/N$.

If n units are selected by SRSWOR, the total number of possible samples are ${}^{N}C_{n}$.

So the probability of selecting any one of these samples is $\frac{1}{{}^{N}C_{n}}$.

Simple random sampling with replacement (SRSWR): SRSWR is a method of selection of n units out of the N units one by one such that at each stage of selection each unit has equal chance of being selected, i.e., $1/N$.

When n units are selected with SRSWR, the total number of possible samples are N^n.

The Probability of drawing a sample is $\frac{1}{N^n}$.

Ex.1: A Population comprise the following units a, b, c, d, e. Draw all possible samples of size three without replacement.

Sol.1: Since in this case, sample size $(n) = 3$ and population size $(N) = 5$. The total number of possible samples without replacement $= {}^{5}C_{3} = 10$. These are abc, abd, abe, acd, ace, ade, bcd, bce, bde, cde.

Ex.2: A Population comprise 3 member $1, 5, 3$. Draw all possible samples of size two with replacement and without replacement.

Sol.2: Since in this case, sample size $(n) = 2$ and population size $(N) = 3$. The total number of possible samples without replacement $= {}^{3}C_{2} = 3$. These are (1,3), (1,5), (3,5).

The total number of possible samples with replacement $= 3^2 = 9$. These are (1,1), (1,5), (1,3), (5,1), (5,5), (5,3), (3,1), (3,5), (3,3).

Random sampling may be done either by adopting lottery method or by using a table of

random number.

Random sampling is used in experimental medicine or clinical trials like testing the efficacy of a particular drug. For, example, Suppose we want to put 20 patients on a clinical trial out of 120 available. To ensure randomness of selection of 20 patients we may adopt lottery method or random number table.

Lottery Method The lottery method is practicable when the population size is comparatively small. To select a random sample by this method, first assign serial numbers as $1, 2, 3, ...$ to each item. Write these numbers on pieces of paper of equal size and of the same quality. After this, roll the papers called 'lots' and shuffle them thoroughly. Take one lot at random. Note the number on it. Now select the item corresponding to this number into the sample. Repeat the process till we get the required number of items into the sample.

Random Number Method This method is adopted with the aid of random number tables. These are table of numbers in which digits selected by a mechanical process of randomization are tabulated. They are tabled as 2 digital, $3, 4$ or 5 digital numbers. After assigning serial numbers to each sampling unit, open any page of the random number table. Then a blind-fold selection of a number is made. Starting from that number we can proceed along a row, column, or diagonally the successive numbers there occur will give a random selection of items. This method can be used to select a random sample even when the population is large.

Stratified Random Sampling: Stratified random sampling is generally used when population is heterogeneous. In this method of sampling the entire population is divided into certain homogeneous subgroups depending upon the characteristic to be studied and random samples are drawn independently from each of the subgroups. These subgroups are known as 'strata'.

As an example, if it is known that the prevalence of a certain disease is different in different age groups then to estimate the prevalence rate of the disease from the sample, the survey is carried out taking stratified random sample from each of the age groups of he population, as the number of people in each age group is not equal in the population.

If proper stratification can be made so that the strata differ form each other as much as possible, but there is much homogeneity within each stratum, than a stratified sample with give better estimates of the population characteristics than a random sample of the same size.

Systematic Random Sampling: This a technique of forming a sample in some systematic manner, usually by taking items at regular intervals. If the population size is finite, all the units are first serially listed and arranged in order. Then from the first r^{th} units, one unit is selected at random. This unit and every r^{th} unit of the serially listed population combined together constitute a systematic sample. r refers to the sample interval which is calculated by the formula:

$$r(\text{sample interval}) = \frac{\text{Total population}}{\text{Sample size desired}}.$$

e.g. if 10% sample is to be taken out of 1000 i.e. $r = \frac{1000}{100} = 10$

A number between 1 to 10 say 4, is selected at random with the help of lottery method by putting out one card after shuffling, out of 10 cards serially numbered 1 to 10. Then the sample will consist of units with serial number $4, 4+10=14, 14+10=24, \ldots\ldots\ldots$ so In the sample, the sample units are 4, 10, 24, 34, 44, 54, 64, 74, 84, 94,..... This sampling is similar to that of selecting every alternative patient or to select every fourth or fifth patient who attends the clinic, after selecting the first patient at random. In this case nothing will be known about the selected patients in advance.

Multistage Sampling In this type of sampling, sampling units are selected at various stages. As as illustration, if the prevalence of a disease is to be estimated in an area, first a sample of village may be selected at random in the first stage, and out of these selected village a random sample of house can be selected in the second stage and out of these sampled house a random of individuals may be selected in the third stage. The sampling designs may be either same or different at each stage.

Non Random Sampling: In non random sampling, the samples are drawn without following any specific procedure or any yardstick. The sample collected does not show any specific approach. In this sampling procedure many investigator biases are likely to occur.

Types of Non Random Sampling The nonrandom sampling can be of following types:

Purposive sampling or Judgement sampling: In this sampling, the individual units of a sample are chosen as according to convenience and personal choice of the investigator. Purposive selection is always subject to some kind of bias. This method is suitable for small sample.

Quota Sampling: In this sampling, the investigators are definite quotas according to certain criteria. The investigators select the individuals i.e. sampling units for interview on their personal judgements within the quotas.

References

- S.C. Gupta, V.K. Kapoor, Fundamentals of Applied Statistics, Sultan Chand & Sons.
- A. Kumar, A. Chaudhary, Survey Sampling, Krishna's Educational Publishers.
- K.S. Negi, Biostatistics, Aitbs Publishers.
- V.B. Rastogi, Fundamentals of Biostatistics, ANE Books.