8. Sampling ~~Theory~~ theory

8.1 Meaning and objects of sampling :

Sampling denotes the selection of a part of the aggregate statistical material with a view to obtain information about the whole. This aggregate or totality of statistical information on a particular character of all the members covered by an investigation, is called Population or Universe. The selected part, which is used to ascertain the characteristics of population is called sample. While taking a sample, the population is assumed to be composed of individual units or members, some of which are included in the sample. The total number of members of the population and the number included in the sample, are called population size and sample size respectively.

A statistical ~~population~~ population is finite or infinite according to its size. When the number of members of the population can be expressed as a definite quantity, it is said to be finite. Otherwise, the population is infinite. In particular, if a ~~sam~~ sample is drawn from a continuous ~~pro~~ probability distribution, then the ~~population~~ population is infinite. ~~As complete sight~~ Again the population may be existent or hypothetical. The population of income of all persons in a country is an example of a existent population. The population of points obtained in all possible throws of a die is an example of hypothetical population.

If a sample of adequate size is properly chosen and analyzed, it is most likely to ~~reveal~~ reveal the characteristics of the whole population and the results obtained therefore can be fairly relied upon as if

they were based on all members of the population. The possibility of reaching valid conclusions concerning a population by means of a properly chosen sample is based on two important laws: (i) Law of Statistical Regularity and (ii) Law of Inertia of Large Numbers. The former states that a sample of reasonably large size when selected at random, is almost sure to represent the characteristics of the population. The second law states that samples of large size show a high degree of stability, i.e., the results obtained from them are expected to be very close to the population characteristics.

8.2 Some ideas about the methods of selecting samples

The two basic principles of sampling are validity and optimization.

By validity of a sampling we mean that the sample should be so selected that the results obtained from it can be interpreted objectively in terms of probability. Validity is ensured by a random (probability) sample so that each of the population members has a definite preassigned probability of being included in the sample.

The principle of optimization takes into consideration the factors of efficiency and cost. Efficiency is measured by the inverse of the sampling variance of the estimator (which is a function of sample values used for guessing about an unknown population characteristic) and cost is measured by money or man-hours spent in the whole process. The principle of optimization is satisfied if a given level of efficiency is reached with minimum cost or maximum efficiency is attained with a given level of cost.

Conducting a sampling involves three stages. They are planning stage, execution stage and analysis or reporting stage.

In the planning stage, you have to

(i) define the purpose of sampling
(ii) define the population
(iii) decide on the nature of the data.
(iv) select method of collecting data
(v) select the sampling unit
(vi) train the personnel needed

In the execution stage the sampled members in the field are identified and the questionnaires are suitably filled.

In the analysis and reporting stage, you have to

(i) scrutinize the data
(ii) tabulate the data
(iii) analyse the data
(iv) report and store the information for future sampling.

A survey conducted on a suitable sample is called sample survey, while in complete enumeration or complete census, the entire population is surveyed.

Generally sample survey is preferred to complete enumeration for cost reduction, greater speed, greater accuracy etc.

8.3 Statistic and parameter.

Any statistical measure calculated on the basis of sample observations is called a Statistic; eg., sample mean, sample standard deviation etc. Any statistical measure based on all units in the population is called a parameter; e.g., population mean, population standard deviation etc. the value of a statistic varies from sample to sample but parameter always is the same.

Usually parameters are unknown and statistics are used as estimates of parameters. The probability distribution of a statistic is called the sampling distribution.

8.4 Four fundamental distributions derived from the normal distribution :

(i) Standard Normal distribution : If a random variable x is normally distributed with mean μ and standard deviation σ then $z = \frac{x-\mu}{\sigma}$ is called a Standard Normal Variate. The probability distribution of z is called Standard Normal distribution and is defined by the p.d.f

$$p(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \quad (-\infty < z < \infty)$$

Characteristics : 1. The standard normal distribution is a special case of normal distribution with mean = 0 and s.d (standard deviation) = 1

2. It has no parameters (unlike normal distribution which has two parameters μ and σ).

3. The central moments are $\mu_2 = 1$, $\mu_3 = 0$, $\mu_4 = 3$

Also, $\beta_1 = 0$, $\beta_2 = 3$; Skewness $(\gamma_1) = 0$, Kurtosis $(\gamma_2) = 0$

4. The standard normal curve is symmetrical about mean 0 and two tails of the curve extends to infinity on either side of the mean. The points of inflection are $z = \pm 1$.

5. The following results regarding the area under the curve are specially useful : (i) Area between $z = \pm 3$ is 99.73%

(ii) Area between $z = \pm 2.58$ is 99% (iii) Area between $z = \pm 1.96$ is 95%

Theorem : If $\bar{x}$ denotes the mean of a random sample of size n from a normal population with mean μ and s.d. σ,

Then $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ follows standard Normal distribution,