13B. SAMPLING

CONCEPT 01 : INTRODUCTION TO SAMPLING

Population (Universe)

- It can be defined as the aggregate of all the units under consideration.
  E.g. Population of students enrolled for CA Foundation.
- The No. of Units belonging to a population is known as **Population Size (N)**.
- The study of every element of population is called **Census**.

TYPES OF POPULATION

Type		*Example*
Finite Population	Population containing finite no. of units.	*Population of students enrolled for CA Foundation*
Infinite Population	Population containing infinite or uncountable no. of units.	*Population of Stars, Mosquitos, Flowers, Insects*
Existent Population	Population consisting of real objects	*Population of a town*
Imaginary Population	Population that exists hypothetically	*Population of heads of a coin tossed infinitely.*

Sample

- It can be defined as a part of population so selected with a view to represent the population in all its characteristics.
- The units forming sample are known as **Sampling Units**.
- A detailed & complete list of all the Sampling Units is known as **Sampling Frame**.
- If a sample contains 'n' units, then 'n' is known as **Sample Size**.

Parameter

- It can be defined as a **characteristic of a population based on all the units of the population**.
- Statistical Inferences are drawn about population parameters based on the sample observations drawn from that population.

Statistic (T)

- It can be defined as a **statistical measure of sample observation** & it is a function of sample observations.

CONCEPT 02 : SAMPLE SURVEY

Meaning

It is the study of the unknown population on the basis of a proper representative sample drawn from it.

Principles of Sample Survey

Law of Statistical Regularity	If a sample of fairly large size is drawn from the population at random, then on an average the sample would possess the characteristics of that population.
Principle of Inertia	The results are more reliable, accurate & precise as the sample size increases, provided other factors are kept constant.
Principle of Optimisation	An optimum level of efficiency at a minimum cost or the maximum efficiency at a given level of cost can be achieved with the selection of an appropriate sampling design.
Principle of Validity	A sampling design is valid only if it is possible to obtain valid estimates & valid tests about population parameters (Only Probability Sampling).

Errors / Bias in Sample Survey

It can be defined as the deviation between the value of population parameter as obtained from a sample & its observed value.

TYPES OF ERRORS

[1] SAMPLING ERRORS

- Since only a part of the population is investigated in a sampling, every sampling design is subject to this type of errors.
- **Factors**
  (a) Due to Defective Sampling Design
  (b) Due to Substitution (of a Sampling Unit)
  (c) Due to Faulty Demarcation of Units
  (d) Due to wrong choice of Statistic
  (e) Variability in Population

[2] NON-SAMPLING ERROR

- This type of error happens in both Sampling & Complete Enumeration.
- **Factors :** Lapse of Memory, Preference for certain digits, ignorance, psychological factors like vanity, non-response on part of interviewees, wrong measurement of sampling units, communication gap, incomplete coverage.

CONCEPT 03 : TYPES OF SAMPLING

Types	PROBABILITY SAMPLING	NON PROBABILITY SAMPLING	MIXED SAMPLING
	1. Simple Random Sampling 2. Stratified Sampling 3. Multi-Stage Sampling	Purposive / Judgment Sampling	Systematic Sampling

Probability Sampling

In this there is always a fixed, pre-assigned probability for each member of the population to be a part of the sample taken from that population.

Simple Random Sampling

- Each unit of sample has an equal chance of being selected.
- It is very simple & effective method provided :
  (a) The population is not very large.
  (b) The sample size is not very small.
  (c) The population under consideration is not heterogeneous.
- It is completely free from Sampler's Bias.
- All the tests of significance are based on the concept of Simple Random Sampling.

Stratified Sampling

- If the population is large & heterogeneous, we divide them into a number of sub-populations (strata) in such a way that there should be little variations among units in a stratum & maximum difference among different strata.
- If Simple Random Sampling is applied for drawing units from all strata, it is known as **Stratified Random Sampling**.
- Purpose of Stratified Sampling are :
  (a) To make representation of all the sub-populations
  (b) To provide an estimate of parameter not only for all the strata but also an overall estimate
  (c) Reduction in variability and thereby an increase in precision
- **Types of Allocation of Sample Size**
  1. **Bowley's Allocation (Proportional Allocation)**
     When there is prior information that there is not much variation between the strata variances, we use Bowley's Allocation, where the sample sizes for different strata are taken proportional to the population sizes i.e. $n_i \mu N_i$.
  2. **Neyman's Allocation**
     When the strata variances differ significantly among themselves, we use Neyman's Allocation, where sample size vary jointly with population size & population standard deviation i.e. $n_i \mu N_i S_i$.

  Here,
  → n_i = Sample Size for i^{th} stratum,
  → N_i = Population Size
  → S_i = Population Standard Deviation
- It is not advisable if :
  (a) Population is not large.
  (b) Some prior information is not available.
  (c) There is not much heterogeneity among the units of population.

Multi Stage Sampling

- In this type of sampling design, sampling is carried out through stages.
  → Firstly, only a number of first stage units are selected.

	→ For each of the selected first stage sampling units, a number of second stage sampling units are selected. → The process is carried on until we select the ultimate sampling units. ▪ The coverage is very large. ▪ It also saves computational labour and is cost effective. ▪ It also adds flexibility in sampling process which is lacking in other sampling schemes. ▪ It is less accurate than stratified sampling.
Non-Probability Sampling	No Probability is attached to the member of the population and as such it is based entirely on the judgment of the sampler. ***Purposive / Judgment Sampling*** ▪ It is dependent solely on the discretion of the sampler & he applies his own judgment based on his beliefs, prejudice, whims, and interest to select the sample. ▪ It is purely subjective & no statistical hypothesis can be tested on the basis of this.
Mixed Sampling	It is based partly on some probabilistic law & partly on some pre-decided rule. ***Systematic Sampling*** ▪ It refers to a sampling scheme where the sampling units are selected at regular interval after selecting the first unit at random (with equal probability). ▪ **Linear Systematic Sampling :** If N is a multiple of 'n', then N = nk, where 0 < k < n, then we are selecting first unit at random from the first k units & thereby selecting every k^{th} unit till the complete, adequate & updated sampling frame comprising all the members of the population is exhausted. ▪ k = **Sample Interval**. ▪ **Circular Systematic Sampling :** If N is not a multiple of 'n', then N = nk + p, p<k & then we select the first unit from first k units at random & thereafter selecting every k^{th} unit in a cyclic order. ▪ It is a very convenient method where a complete & updated Sampling Frame is available. ▪ It is less time consuming, less expensive & simple as compared to other methods of sampling. ▪ If there is an unknown & undetected periodicity in the sampling frame & sampling interval is a multiple of that period, then we are going to get a most biased sample. ▪ Since, it is not Probability Sampling, no statistical inference can be drawn about population parameter.

CONCEPT 04 : SAMPLING FLUCTUATION & SAMPLING DISTRIBUTION	
Sampling Fluctuation	The variation in value of a statistic computed from different samples.
Sampling Distribution	▪ The Probability Distribution of a given statistic is known as **Sampling Distribution**. ▪ The Mean of a Statistic, as obtained from its Sampling Distribution, is known as **Expectation**. ▪ The Standard Deviation of the Statistic is known as the **Standard Error (SE)**. ▪ Standard Error (SE) can be regarded as a measure of precision achieved by sampling. ▪ Standard Error (SE) is inversely proportional to the $\sqrt{\text{Sample Size (n)}}$. ▪ Starting with a population of N units, we can draw many a sample of a fixed size 'n'. **(a) Total No. of Samples (With Replacement) =** N^n **(b) Total No. of Samples (Without Replacement) =** ${}^{N}C_{n}$