Sampling and Testing of Hypothesis

Some Basic Definitions

Population (or Universe):

The totality of statistical information on a particular character, from all members covered by an enquiry is called a population or universe. It may be finite or infinite. Its size is denoted by *N*.

Sample:

A selected part of the population is called a sample. It throws light on the population characteristics. Its size is denoted by *n* (< *N*).

[Normally $n \geq 30$ ($n < 30$) signifies a large (small) sample]

Census:

The complete enumeration of a population at a point of time with respect to some well-defined characteristics is called the census.

Sampling Survey:

A survey of a part of population is called sampling survey.

Population Parameter:

Any statistical measure computed from the population is called a population parameter. For example: population mean (μ), population variance (σ^2) etc. The parameter is usually denoted by θ.

Sample Statistic:

Any statistical measure computed from the sample observations will be called a sample statistic. For example: sample mean ($\bar{x}$), sample variance (s^2) etc. The statistic is usually denoted by *t*.

Sampling Fluctuation:

The differences in the values of sample statistic for different samples is called sampling fluctuations.

Estimation:

It is a rule or method of estimating a population parameter by using a sample statistic. It is generally expressed as a function of sample observations.

Unbiased Estimator:

An estimator *t* is called an unbiased estimator of a population parameter θ if $E(t) = \theta$. Otherwise *t* is called a biased estimator.

Sampling Distribution:

The probability distribution of all possible samples of a certain size drawn from a particular population is called the sampling distribution.

Sampling Error:

The difference between population parameters and sample statistic is called sampling error. For example, $\mu \sim \bar{x}$ or $\sigma^2 \sim s^2$. This is an unavoidable error and inherent but this can be reduced by increasing sample size.

Standard Error:

The standard deviation calculated from the sampling distribution of a test statistic is called the standard error. It is considered to be a measure of sampling error. It decreases with the increase in the sample size.

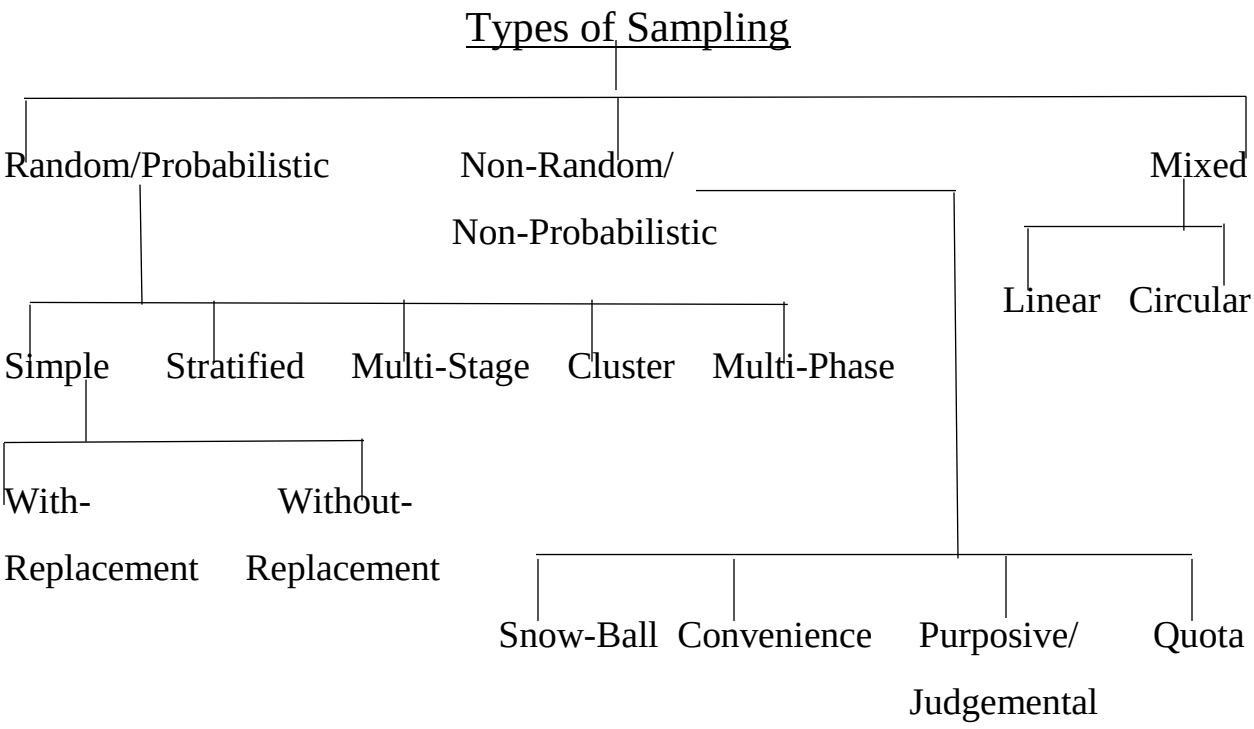
Non-Sampling Error:

These errors occur due to (i) faulty measurements

(ii) observational mistakes

(iii) incorrect editing/coding/tabulation

They are likely to increase with increase in sample size. Due to this characteristics, they are sometimes known as cumulative errors.



Note:

(i) Since statistic *t* varies from sample to sample, it is a random variable and thus has a probability distribution

(ii) For Simple Random Sampling With Replacement, if population size = N and sample size = n, then number of all possible samples = N^n and probability of drawing each sample = $\frac{1}{N^n}$

(iii) For Simple Random Sampling Without Replacement, if population size = N and sample size = n, then number of all possible samples = ${}^{N}C_n$ and probability of drawing each sample = $\frac{1}{{}^{N}C_n}$