UNIT-III

* Sampling Distribution *

Population :- The totality of observation with which we are concern, whether this number be finite or infinite is called Population.

(or)

Population is the agregate or totality of statistical data, a subject of investigation.

For example, the population of the heights of Indians.

the population of nationalized banks in India.

Size of the population:-

The no. of observations in the population is defined to be the size of the population. It may be finite or infinite and size of the population (or) is denoted by N.

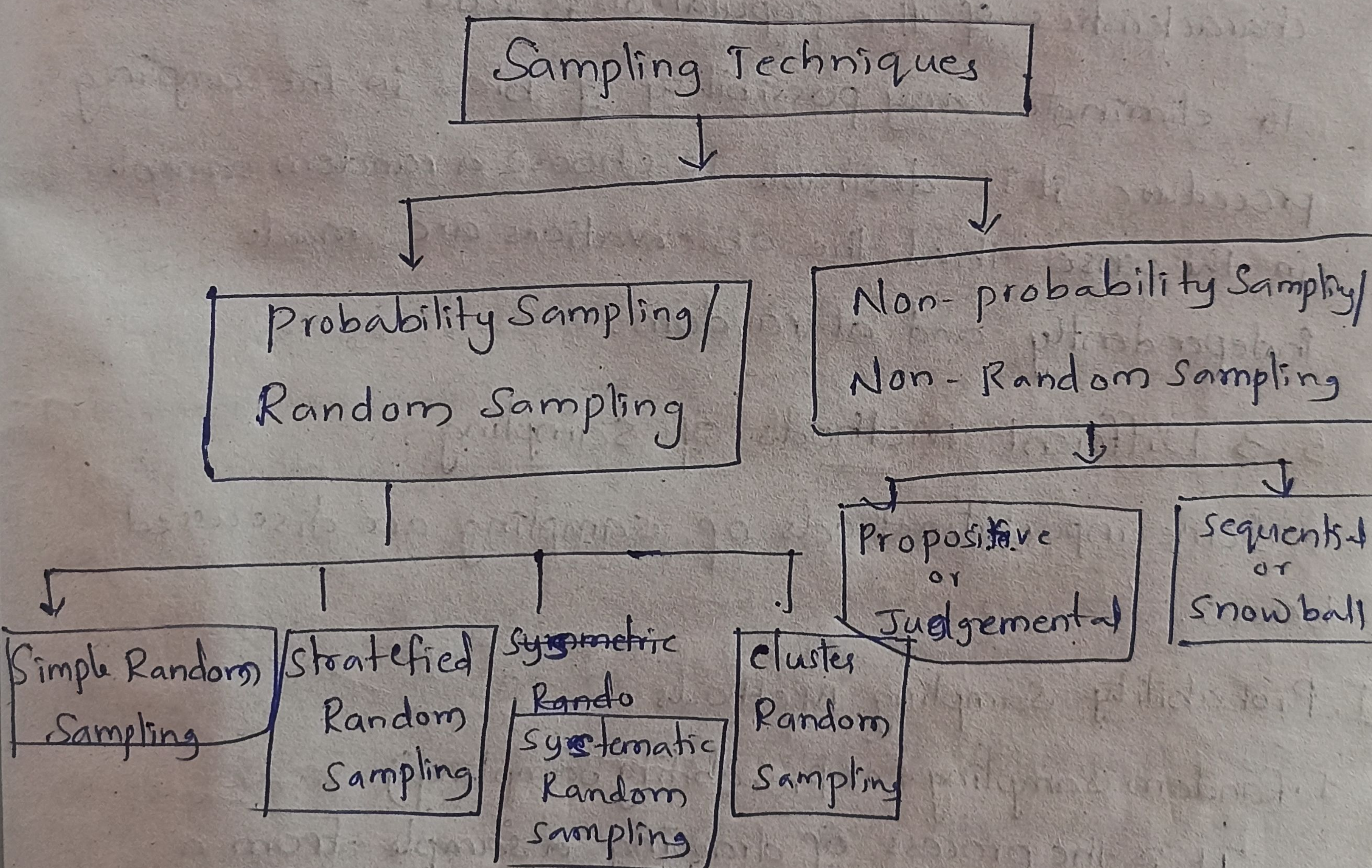
Ex:- If there are 60 students in a college, that we classified according to blood groups we say that we have a population of size 600.

Sample or Sampling:-

Sample is defined as the subset of population is known as sample and it is denoted by 'n'.

① Difference b/w Sampling and Non-sampling methods (or)

Diff. b/w probabilistic and Non-probabilistic techniques.



Sampling Distributions:-

The population is thus a universal set for the sample. The statistical constants like mean, standard deviation correlation coefficient, etc.. obtained for the population are called parameters.

Similarly, constants for the sample drawn from the given population i.e; mean ($\bar{x}$), standard deviation (S) etc. are called the statistic.

If our inferences from the sample to the population are to be valid, we must obtain samples that are representative of the population. All too often, we are tempted to choose a sample by selecting the most convenient members of the population. Such a procedure may lead to erroneous inferences concerning the population. Any sampling procedure that produces inferences that consistently over estimate (or) consistently under estimate some characteristics of the population is said to be biased. To eliminate any possibility of bias in the sampling procedure it is desirable to choose a random sample in the sense that the observations are made independently and at random.

5.3 Different Methods of Sampling.

Some important methods of sampling are discussed below.

I. Probability Sampling Methods.

1. Random Sampling (or Probability Sampling)

It is the process of drawing a sample from a population in such a way that each member of the population has an equal chance of being included in the sample. The sample obtained by the process of random sampling is called a random sample.

For example: (i) A handy of cards dealt from a well-shuffled pack of cards is a random sample

(ii) Selecting randomly 20 words from a dictionary is a random sample.

(iii) Choosing 10 patients from a hospital in order to test the efficacy of a certain newly-invented drug.

If each element of a population may be selected more than once then it is called sampling with replacement whereas if the element cannot be selected more than once, it is called sampling without replacement.

Note:- If N is the size of a population and n is the sample size, then

(i) The number of samples with replacement $= N^n$

(ii) The number of samples without replacement $= {}^{N}C_{n}$.

2. Stratified Sampling (or Stratified Random Sampling)

The method is useful when the population is heterogeneous. In this type of sampling, the population is first sub-divided into several parts (or small groups) called strata according to some relevant characteristics so that each stratum is more or less homogeneous. Each stratum is called a sub-population. Then a small sample (called sub-sample) is selected from each stratum at random. All the sub-samples are combined together to form the stratified sample which represents the population properly. The process of obtaining and examining a stratified sample with a view to estimating the characteristic of the population is known as Stratified Sampling.

For example, let us select a stratified sample of 500 families from a city having 50,000 families, with a view to studying their economic condition. For this purpose, the city area is divided into a number of strata, according to economic condition of their inhabitants, as measured by manual income (say). Thus, localities mostly inhabited by people with more or less similar annual income (say). Thus, localities mostly inhabited A few families are then chosen at random from each so that the sum total of all the families from all the strata is 500.

3. Systematic Sampling (or Quasi- Random Sampling)

As the name suggests this means forming the sample in some systematic manner by taking items at regular intervals. In this method, all the units of the population are arranged in some order. If the population size is finite, all the units of the population are arranged in some order. Then from the first k items, one unit is selected at random. This unit and every kth unit of the serially listed population combined together constitute a systematic sample. This type of sampling is known as systematic sampling. The difference between random sampling and systematic sampling lies in the fact that in the case of a random sample all the members have to be chosen randomly, whereas in the case of a systematic sample only the first member has to be chosen at random.

II. Non- Probability Sampling Methods:-

4. Purposive Sampling (or Judgement Sampling).

When the choice of the individual items of a sample entirely depends on the individual judgement of the investigator (or sample), it is called a Purposive or Judgement sampling.

In this method, the members constituting the sample are chosen not according to some ~~definite~~ definite scientific procedure, but according to convenience and personal choice of the individual, who selects the sample. Two or more such independent purposive samples may give widely different estimates of the same population. In this type, the investigator must have a good deal of experience and a through knowledge of the population. Purposive selection is always subject to some kind of bias. This method is suitable when the sample is small.

For example, if a sample of 20 students is to be selected from a class of 100 to analyse the extra-curricular activities of the students, the investigator would select 20 students who, in his jubgement, would represent the class.

5. <u>Sequential Sampling</u>:-

It consists of a sequence of sample drawn one after another from the population depending on the results of previous samples. If the result of the first sample leads to a decision which is not acceptable, the lot from which the sample was drawn is rejected. But if the result of the first sample is acceptable, no new sample is drawn. But if the first sample leads to no clear decision, a second sample is drawn and, as before, if required, a third sample is drawn to arrive at a final decision to accept or reject the lot. It is widely used in statistical Quality Control in factories engaged in mass production and other areas.

02/01/2025

Central limit theorem:-

If $\bar{x}$ with the mean of random sample of size 'n' drawn from a population having mean (μ) and standard deviation (σ) then the sampling distribution $\bar{x}$ or Standard Error (S.E) = $\frac{\sigma}{\sqrt{n}}$ provided the sample size n is large $(n \geq 30)$

$$\therefore Z = \frac{\bar{x} - \mu}{S.E} \quad \text{or} \quad \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

where $\bar{x}$ = sample mean.

μ = population mean.

σ = population S.D

n = sample size.

Large Sample:-

If sample size $n \geq 30$ be considered as large sample.

Small Sample:-

If the sample size $n < 30$ is known as small sample.

Problems on Sampling Distribution.

*****

(1) A populatn consist a five members 2, 3, 6, 8 & 11. consider all possible samples of size 2 which can be drawn with replacement from this populatn. Find.

(i) The mean of the population

(ii) The S.D of the population.

(iii) The mean of the sampling Distribution of means.

(iv) The standard Error of means.

(i) Mean of population (μ).

$$\bar{x} \text{ or } \bar{\mu} = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5}{5}$$

$$= \frac{2+3+6+8+11}{5} = 6.$$

(ii) population S.D (σ).

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} = \sqrt{\frac{(2-6)^2+(3-6)^2+(6-6)^2+(8-6)^2+(11-6)^2}{5}}$$

$$\sigma = \sqrt{\frac{16+9+0+4+25}{5}} = \sqrt{\frac{54}{5}}$$

$$\sigma = \sqrt{10.8} = 3.286$$

(iii) Mean sampling Distribution of Means with Replacement $N^n = 5^2 = 25$

$$\left.\begin{array}{ccccc}
(2,2) & (2,3) & (2,6) & (2,8) & (2,11) \\
(3,2) & (3,3) & (3,6) & (3,8) & (3,11) \\
(6,2) & (6,3) & (6,6) & (6,8) & (6,11) \\
(8,2) & (8,3) & (8,6) & (8,8) & (8,11) \\
(11,2) & (11,3) & (11,6) & (11,8) & (11,11)
\end{array}\right\} \text{25 ways}$$

Mean of samplings.

$$\begin{bmatrix}
\frac{2+2}{2} = 2 & 2.5 & 4 & 5 & 6.5 \\
2.5 & 3 & 4.5 & 5.5 & 7 \\
4 & 4.5 & 6 & 7 & 8.5 \\
5 & 5.5 & 7 & 8 & 9.5 \\
6.5 & 7 & 8.5 & 9.5 & 11
\end{bmatrix} = 150$$

$$\mu_{\bar{x}} = \frac{\text{Sampling Distribution of Mean}}{25}$$

$$= \frac{150}{25} \quad \frac{2 + 2.5 + 4 + 5 + 6.5 + \cdots + 9.5 + 11}{25}$$

$$= \frac{150}{25} = 6$$

$$\therefore \bar{\mu} = \mu_{\bar{x}}$$

(iv) $S.E = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{10.8}}{\sqrt{2}} = \frac{\cancel{319.8}}{1.414} = \cancel{7.636}$

$= \frac{3.28}{\sqrt{2}} = 2.31$

$$\boxed{S.E = 2.31}$$

*****

(2) If the populat^n is 3, 6, 9, 15, 27. Find (Sample size is 3)

(i) Mean of the population.

(ii) S.D of population.

(iii) Mean of sampling distribution

(iv) Standard Error.

(i) $\bar{\mu}$ or $\bar{x} = \frac{\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5}{5}$

$= \frac{3+6+9+15+27}{5} = \frac{60}{5} = 12.$

$\bar{\mu}$ or $\bar{x} = 12.$

(ii) $\sigma = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n}} = \sqrt{\frac{(3-12)^2 + (6-12)^2 + (9-12)^2 + (15-12)^2 + (27-12)^2}{5}}$

$= \sqrt{\frac{360}{5}} = \sqrt{72} = 8.485.$

(iii) Mean sampling Distribution of means.

without Replacement $^{N}C_n = {}^{5}C_3 = 10$

(3, 6, 9, 15, 27).

(3, 6, 9) (3, 6, 15) (3, 6, 27) (3, 9, 15) (3, 9, 27) (3, 15, 27)

(6, 9, 15) (6, 9, 27) (6, 15, 27)

(9, 15, 27)

} 10 ways

Mean of S. Distribution

$$\begin{bmatrix} \frac{3+6+9}{3} = 6, & 8, & 12, & 9, & 13, & 15 \\ 10 & 14 & 16 \\ 17 \end{bmatrix}$$

$$\mu_{\bar{x}} = \frac{\text{Sampling Distribution of Mean}}{10}$$

$$= \frac{6+8+12+9+13+15+10+14+16+17}{10}$$

$$\mu_{\bar{x}} = \frac{120}{10} = 12$$

$$\therefore \bar{\mu} = \mu_{\bar{x}}$$

(iv) $S.E = \frac{\sigma}{\sqrt{n}} \times$ correction factor.

$$\left[\frac{N-n}{N-1}\right] \times \frac{\sigma}{\sqrt{n}} = \left(\frac{5-3}{5-1}\right)\left[\frac{8.485}{\sqrt{3}}\right]$$

$$= 0.5(4.89) = 2.445$$

$$\boxed{S.E = 2.445}$$

3,15,27) ways 17

08/01/2025.

Test of Hypothesis - I

Definition of Statistical Hypothesis :-

The Decision about the population on the basis of sample informatn, we make the assumptions or guesses about the population parameters involved. Such an assumption is called A Statistical Hypothesis. which may or may not be true.

Types of Hypothesis

1> Null Hypothesis (H_0) :- For applying the tests of significance, the first ~~sector~~ setup hypothesis of no difference is called Null Hypothesis.

$$H_0 : \mu = \mu_0$$

2> Alternative Hypothesis (H_1) :- Any Hypothesis which complementary to the null hypothesis is called an Alternative hypothesis, usually denoted by H_1.

conditions :-

$\mu \neq \mu_0$ (Two tailed test)

$\mu > \mu_0$ (RTT) Right Tailed test } one TT

$\mu < \mu_0$ (LTT) Left " "

Errors of sampling:-

The main objective in sampling theory is to draw valid inferences about the population parameter on the basis of sample results.

Types of errors:-

(i) Type - I error (ii) Type-II error.

(i) Type-I error:- Reject H_0 when it is true.

It is the error of rejecting null Hypothesis H_0 when it is true and it is denoted by α.

(ii) Type II error:- Accept H_0 when it is wrong.

It is the error of accepting the null hypothesis H_0 when it is false and it is denoted by $(1-\alpha)$ or β.

Procedure for Testing of Hypothesis:-

~~Step 1~~:- Null Hypothesis (H_0)

Step 1:- Define or set up a null Hypothesis (H_0) taking into consideration the nature of problem and data involved.

Step 2:- Alternative Hypothesis (H_1)

Setup alternative hypothesis so that we

could decide whether we should use one tailed or two tailed test.

Step 3 :- Level of significance :- Select the appropriate level of significance (α) depending on reliability of the estimates and permissible risks i.e; a suitable α is selected in advance if it is not given in the problem then assume 5% level of significance (LOS).

Step 4 :- Test statistics : Compute the test statistics

$$Z = \frac{t - \text{Expected}(t)}{\text{Standard Error}(t)}.$$

Step 5 :- Conclusion : We compare computed value of the test statistics Z with the critical value or table value Z_α at given level of significance (α).

$\therefore$ 5% = 1.96
10% = 1.645
1% = 2.54

case 1 :- If $|Z_{cal}| < Z_\alpha$

H_0 is accepted

H_1 is rejected.

case 2 :- If $|Z_{cal}| > Z_\alpha$ then

H_0 is rejected

H_1 is accepted.

Methods for Large Sample Test :-

1. Test of Significance for Single Mean :-

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

2. Test of Significance for Difference of Means :-

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(\sigma_1)^2}{n_1} + \frac{(\sigma_2)^2}{n_2}}}$$

3. Test of significance for single proportion :-

$$Z_{cal} = \frac{\hat{P} - P}{\sqrt{\frac{PQ}{n}}}$$

4. Test of Significance for Difference of proportions :-

$$Z_{cal} = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{PQ\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}}$$

Problems of method 1 :-

① The sample of 64 students have a mean weight of 70 kgs can be regarded, as a sample from population with mean weight 56 kgs and S.D 25kgs.

Sol:- Given,

$n = 64$, $\bar{x} = 70$, $\mu = 56$, $\sigma = 25$

$n \geq 30$

$64 \geq 30$.

The given sample is large sample

Given that,

Sample Mean $(\bar{x}) = 70$ kgs.

population Mean $(\mu) = 56$ kgs.

S.D $(\sigma) = 25$ kgs.

Sample size $(n) = 64$ students.

Null Hypothesis

There is no significance difference between sample mean and population mean.

$\mu = \mu_0$

Alternative Hypothesis

There is a significance difference b/w s.m & p.m $\mu \neq \mu_0$ (TTT)

Assume 5% of level of significance.

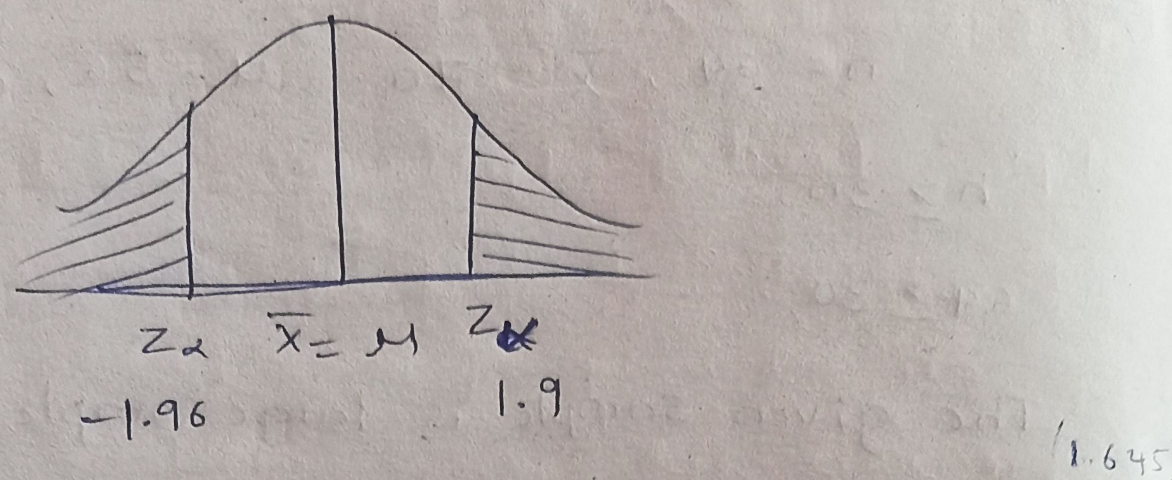
$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{70-56}{25/\sqrt{64}} = \frac{14}{25/8} = \frac{14}{3.125}$$

$Z_{cal} = 4.48$.

$Z_{tab} = 1.96$;

$Z_{cal} > Z_{tab} \Rightarrow H_0$ is rejected, H_1 is accepted

∴ We conclude there is a significance differen between s.m & p.m.



② In a random sample of 60 workers, the average time taken by them to get to work is 33.8 minutes with a standard deviation of 6.1 minutes. Can we reject the null hypothesis ? $\mu = 32.6$ mins in favour of alternative hypothesis $\mu > 32.6$ at $\alpha = 10\%$.

③ A sample of 400 items is taken from a population, the S.D = 10. The mean of the sample is $(\bar{x})$ 40. Test whether the sample has comes from at population with mean 38 (μ). Also calculate 95% conference Interval

Sol:- $C.I = \left[\bar{x} \pm Z_\alpha (\sigma/\sqrt{n}) \right]$

$= 40 \pm 1.96 \left[10/\sqrt{400} \right]$

$= 40 \pm 1.96 \left[\cancel{10}/\cancel{20}_{2} \right]$

$= 40 \pm 1.98 (0.50)$

$= 39.02, 40.98$

$39.02 < 40 < 40.98$

(2)

Sol:- Given ,

$n = 60$

$\bar{x} = 33.8$

$\sigma = 6.1$

$\mu = 32.6$

$n \geq 30 \Rightarrow 60 \geq 30$

H_0:-

There is no significance difference b/w S.M & P.M

$\mu = \mu_0$

H_1:-

There is a significance difference b/w S.M & P.M

$\mu > \mu_0$ (RTT)

We have 10% ~~of~~ level of signifance

$$\boxed{Z_{tab} = 1.645}$$

$$Z_{cal} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{33.8 - 32.6}{6.1/\sqrt{60}} = \frac{1.2}{6.1/7.74}$$

$$Z_{cal} = \frac{1.2}{0.78} = 1.538$$

$$\boxed{Z_{cal} = 1.538}$$

$\therefore \boxed{Z_{cal} < Z_{tab}}$

$\Rightarrow$ H_0 is accepted , H_1 is rejected .

$\therefore$ We conclude there is a significance diff b/w S.M & P.M .