Sampling

By Anurag chauhan

Branches of statistics

Descriptive Stats
⇓
Gathering, classifying summarizing Data from samples

Inferential Stats
⇓
It involves drawing conclusions, generalization or making prediction from the gathered Data

Population

→ Group of People
→ Group of objects
→ Group of events & observations.

Eg Height of male students

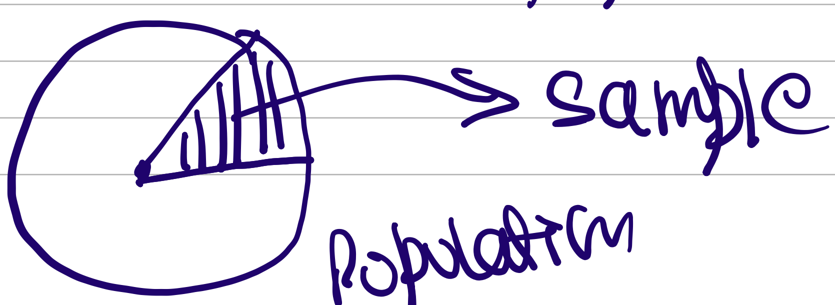
Eg Blood Pressure of females b/w age 40 & 60 years

Eg Temperature in June

Sample

Sample is a subset of Population

" small group of elements selected from populations.



Number of elements in a sample is called the sample size.

Parameters : measurable characters of populations are called parameters

eg Population mean = μ
Population variance = σ^2

Statistics : measurable characters of sample

eg sample mean = $\bar{x}$
sample variance = s^2

Principles of sampling

Sampling is the procedure of selecting elements for a sample from population so that inferences can be drawn about population from sample.

∴ Some Basic Principles of sampling

#1> Law of Statistical Regularity

This law suggests that if a large sample is taken randomly from population, it will possess almost same characters of population.

#2] Law of Inertia of Large Numbers

This law is the corollary

of " Law of statistical Regularity"

This law says that " Larger the size of sample, more accurate the results are "

#3] Principle of optimization

maximum efficiency at minimum cost can be achieved only when appropriate "sampling design" is selected

#4] Principle of validity

According to this law sampling Design is valid only if it is possible to obtain accurate estimates about population

Sampling v/s census

1) Speed : Sampling takes less time

2) cost : Sampling is least expensive

3) Accuracy & Reliablity : Census gives more Accurate & Reliable results unless there is no bias or error in colleting information.

Types of sampling

Probability sampling

non Probability sampling

Probability sampling (Random sampling)

⇓

When elements are Randomly selected for sample

This can be done in 4 ways

i> Simple Random sampling (SRS)

In this sampling each member of population has an equal chance of being selected in sample.

Eg There are 1000 students in a college. You assign a number to every student & then Randomly select a sample of 50 students

ii> Systematic sampling (Quasi Random sampling)

In this sampling every member is of the population is assigned a number. This first member is selected randomly & then instead of chosing other Randomly we chose them in regular intervals

Eg There are 1000 employees in a company, they are assigned numbers from 1 to 1000. we Ranomply select Number '6' from first 10 numbers & after that every 10^{th} person is selected that is 6, 16, 26, 36, ...

iii> Stratified Sampling

In this sampling members are divided into subgroups called strata based on gender, Age & income etc.

After that members are selected using Random or systematic sampling from each sub group.

Eg: let say there are 1000 employees out of which 800 are males & 200 females. A sample of 100 employees reflecting gender balance of the company is made by Dividing the employees in males & females then selecting 80 male & 20 female employees

iv) <u>Cluster sampling (multi Stage sampling)</u>

When population size is large, divide the population in subgroups (each subgroup has similar characteristic of the whole population).

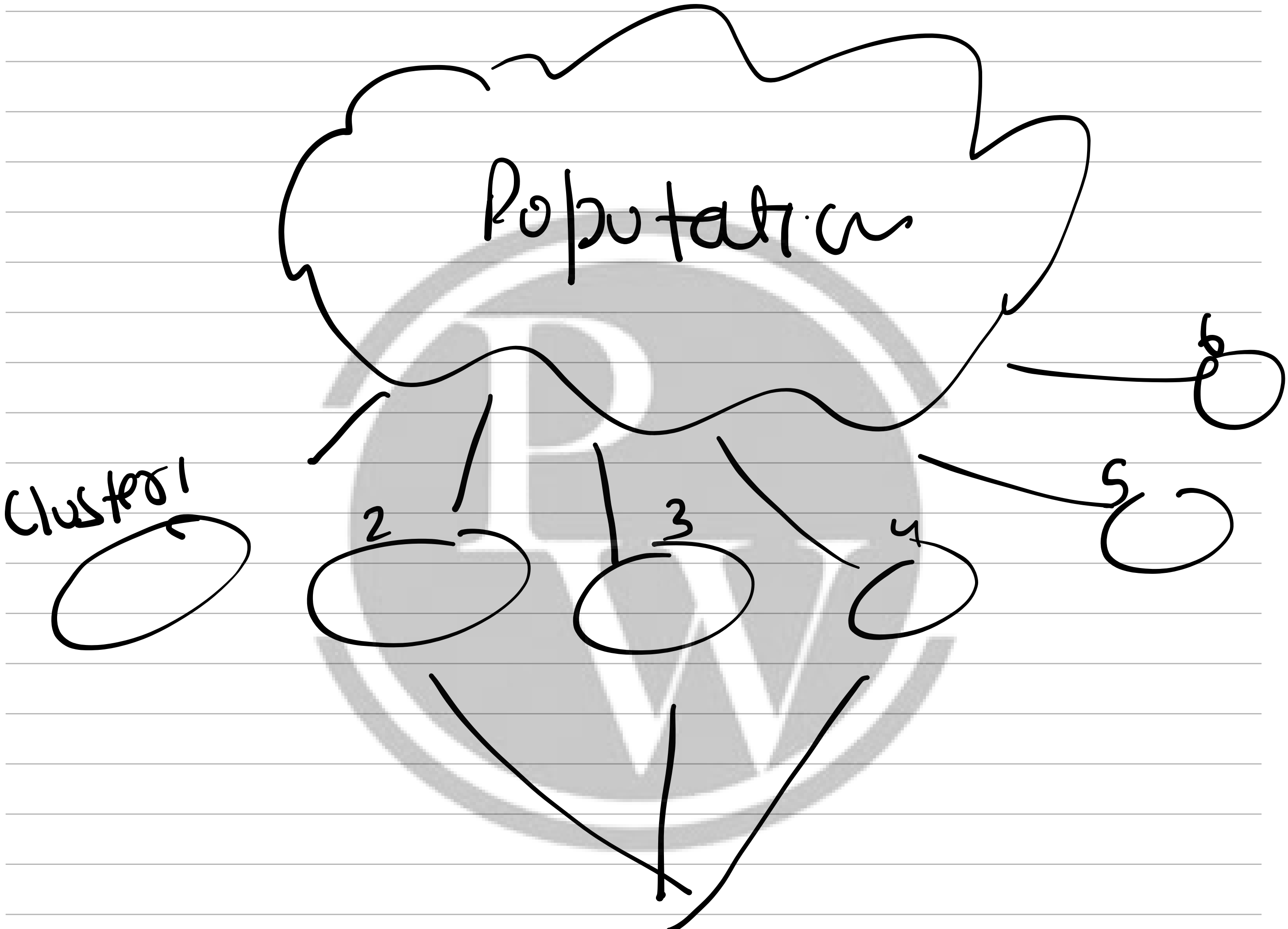
Then some groups are selected randomly & then members are selected from them for sample.

Eg You want to study the behavior of Govt employees of Nagar Nigam.

Nagar Nigam Dept of each state is a cluster. (sample)

Then you select any 3 Department of Nagar Nigam & then select 20 members from each Department.

Thus sample of 60 item is formed.



Any 3 Dept. are selected.

Non Probability sampling

Each member is not selected randomly, so valid inferences can not be made in this type of sampling.

1> Purposive or Judgement sampling

Based on the opinion of expert

eg Indian idol

2> Convenience sampling

Those elements are selected which are easily accessible to researcher.

eg Asking your students to complete survey regarding services provided by universities.

3> volunteer response sampling

People who are themselves ready to conduct the survey collect the sample Data

4> Snow ball Sampling

first select some members, then with the help of them select some more & process continues.

Errors in Sample Survey

Sampling Errors | Non Sampling Errors

Sampling Errors

Difference b/w Sample Statistics & population parameter because sample was not the true representative of population

→ fauty sampling method

→ Faulty Demarcation of sampling units

→ Replacing sampling unit with unsuitable unit

→ wrong choice of statistic

Non sampling errors

These are human errors

census & sampling both can have these errors.

→ Lapse of memory
→ Preference for certain digits
→ wrong measurements
→ untrained interviewer
→ Biased opinion

Population

⇓

Aggregate of all units
under consideration

⇓

Population size
Denoted by 'N'

Finite Population
⇓
finite countable elements

Infinite Population
⇓
Uncountable elements

Population

Existent
⇓
Eg Population of a town
Eg Car Produced by Hyundai

Hypothetical
⇓ (Imaginary)
Eg A coin is tossed 20 times
Eg A card is with replacement infinite times.

⇒ Sample size is denoted by 'n'

⇒ Detailed & complete list of all sampling units is known as sampling frame

Parameter

⇓

Characteristic of Population

⇒ Population mean $(\mu) = \dfrac{\sum_{i=1}^{N} x_i}{N}$

⇒ Population Proportion $(P) = \dfrac{X}{N}$

Eg In a population 10,000 units 1000 units are Defective

then $P = \dfrac{1000}{10,000} = \dfrac{1}{10} = 0.1$

$\Rightarrow$ Population S.D. $(\sigma) = \sqrt{\frac{\Sigma(x_i - \mu)^2}{N}}$

Statistics

measurable characters of sample

$\hat{\mu}$ = Sample mean $= \bar{x} = \frac{\Sigma x_i}{n}$

$\hat{\sigma_x}$ = Sample S.D $= s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$

$\hat{p}$ = Sample Proportion $= p = \frac{x}{n}$

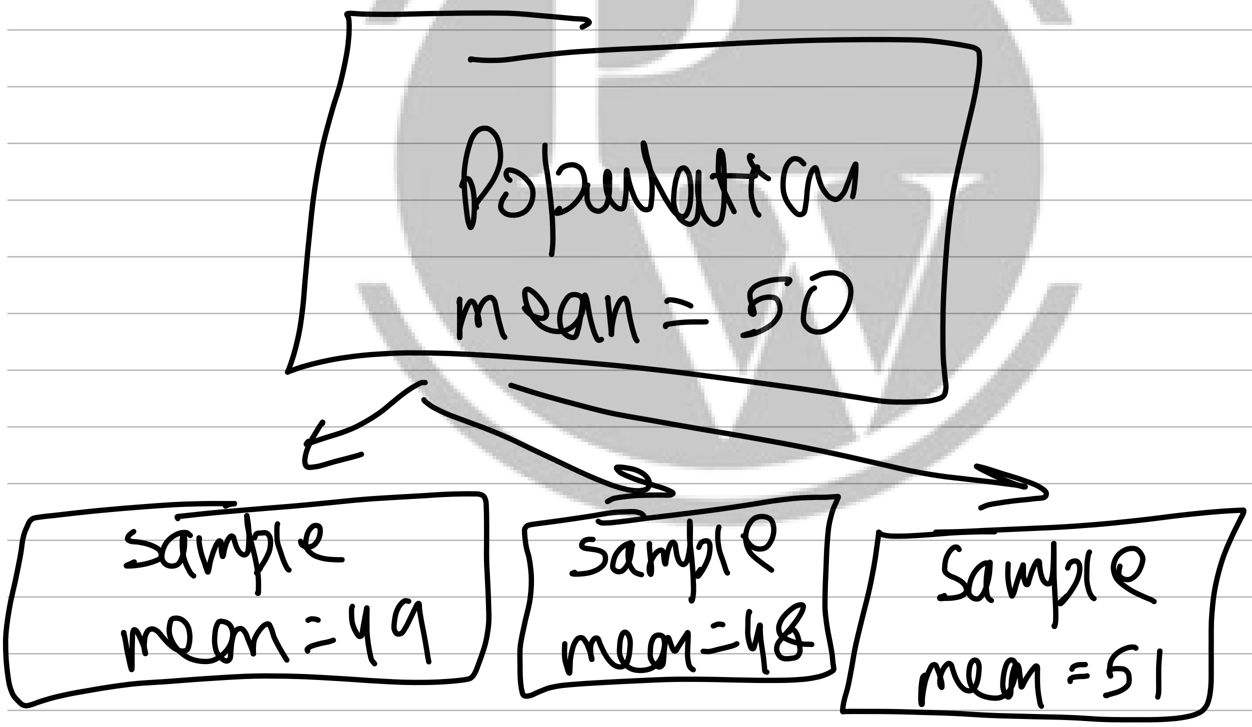
Total no of samples with replacement $= (N)^n$

Total no of samples without replacement $= {}^N C_n$

Sampling fluctuation

⇓

value of sample statistics may be different in different samples, this variation is called sampling fluctuation



in every sample, mean is different, similary SD, correlation also can be different.

This is sampling fluctuations.

Eg A population comprises of following units : a, b, c & d.

Draw all possible samples of

i> Size two without replacement

ii> Size two with Replacement

Sol: i> without Replacement

⇓

element selected once can not be selected again

ab, ac, ad, bc, bd, & cd

Total samples $= {}^{N}C_{n} = {}^{4}C_{2} = \frac{4!}{2!2!} = 6$

ii> with Replacement

⇓

element selected once can be selected again

aa, ab, ac, ad, ba, bb, bc, bd

ca, cb, cc, cd, da, db, dc, dd

Total samples $= N^{n} = 4^{2} = 16$

{Sampling Distribution}

→ we can make many samples of same size with a given population.

→ Sample statistic will be different in each sample.

→ If this sample statistic is considered as a Random variable we can make Probability Distribution

→ This probability Distribution is known as sampling Distribution

[To understand this topic, we must have the conceptual understanding of Prob. Distribution & Random variable concept]

ये बस याद रखना होगा

→ The mean of sampling Distribution is known as Expectation denoted by $E(x)$

→ Standard deviation of this sampling Distribution is known as "Standard Error"

$$\text{Standard Error of mean}(\bar{x}) = \frac{\sigma}{\sqrt{n}} \quad \text{for SRSWR}$$

$$\text{Standard Error of mean} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \quad \text{for SRSWOR}$$

SRSWR → Simple Random sampling with Replacement

SRSWOR → Simple Random sampling without Replacement

Standard Error for Proportion

$$SE(p) = \sqrt{\frac{pq}{n}} \qquad \text{SRSWR}$$

$$SE(p) = \sqrt{\frac{pq}{n}}\sqrt{\frac{N-n}{N-1}} \qquad \text{SRSWOR}$$

$$\sqrt{\frac{N-n}{N-1}} = \text{finite Population multiplier}$$

Q A population has 3 elements 1, 5 & 3.

Draw all possible sample of size two.

i) with Replacement
ii) without Replacement

Also make sampling distribution of sample mean in both cases. (less important for exam)

Sol: i] Population = {1, 5, 3}

sample of two elements

$N = 3$, $n = 2$

Total samples $= N^n = 3^2 = 9$

Sample	Sample mean ($\bar{x}$)
1, 1	1
1, 5	3
1, 3	2
5, 1	3
5, 5	5
5, 3	4
3, 1	2
3, 5	4

3, 3 | 3

For sampling Distribution
sample mean is considered
Random variable
X : Sample mean = 1, 2, 3, 4, 5

x_i	P_i
1	1/9
2	2/9
3	3/9
4	2/9
5	1/9
	1

→ Sampling Distribution

ii) without Replacement
Population {1, 5, 3}
Sample of two elements

Total no of sample = $^{N}C_{n} = {}^{3}C_{2} = 3$

samples	sample mean ($\bar{x}$)
1 & 5	3
1 & 3	2
5 & 3	4

X: Random variable = mean = 2, 3, 4

x_i	p_i
2	1/3
3	1/3
4	1/3
	1

→ Sampling Distribution

Expectation of Sampling Distribution $= E(x) = \sum p_i x_i$

variance of sampling Distribution $= \sum p_i x_i^2 - \left(\sum p_i x_i\right)^2$

Eg Sampling Distribution

x_i	1	2	3
p_i	0.2	0.5	0.3

find mean & variance

Sol:

x_i	p_i	$p_i x_i$	$p_i x_i^2$
1	0.2	0.2	0.2
2	0.5	1.0	2.0
3	0.3	0.9	2.7
	1	2.1	4.9

mean or Expectation $= E(x) = \Sigma p_i x_i$

$= 2.1$

variance $= \Sigma p_i x_i^2 - (\Sigma p_i x_i)^2$

$= 4.9 - (2.1)^2$

$= 4.9 - 4.41$

$= 0.49$

S.D or S.E $= \sqrt{0.49}$

$= 0.7$

Q: In a population of 3 items SD is $\sqrt{\frac{8}{3}}$. If sample of 2 items are made
i> with replacement, find standard error of mean
ii> without replacement, find standard error of mean.

Sol: i> with Replacement

$N=3, \; n=2, \; \sigma = \sqrt{\frac{8}{3}} = \frac{2\sqrt{2}}{\sqrt{3}}$

$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} = \frac{2\sqrt{2}}{\sqrt{3}} \times \frac{1}{\sqrt{2}} = \frac{2}{\sqrt{3}}$

ii> with out Replacement

$N=3, \; n=2, \; \sigma = \sqrt{\frac{8}{3}} = \frac{2\sqrt{2}}{\sqrt{3}}$

$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}} \times \sqrt{\frac{N-n}{N-1}}$

$= \frac{2\sqrt{2}}{\sqrt{3}} \times \frac{1}{\sqrt{2}} \times \sqrt{\frac{3-2}{3-1}}$

$= \frac{2\sqrt{2}}{\sqrt{3}} \times \frac{1}{\sqrt{2}} \times \frac{1}{\sqrt{2}} = \sqrt{\frac{2}{3}}$