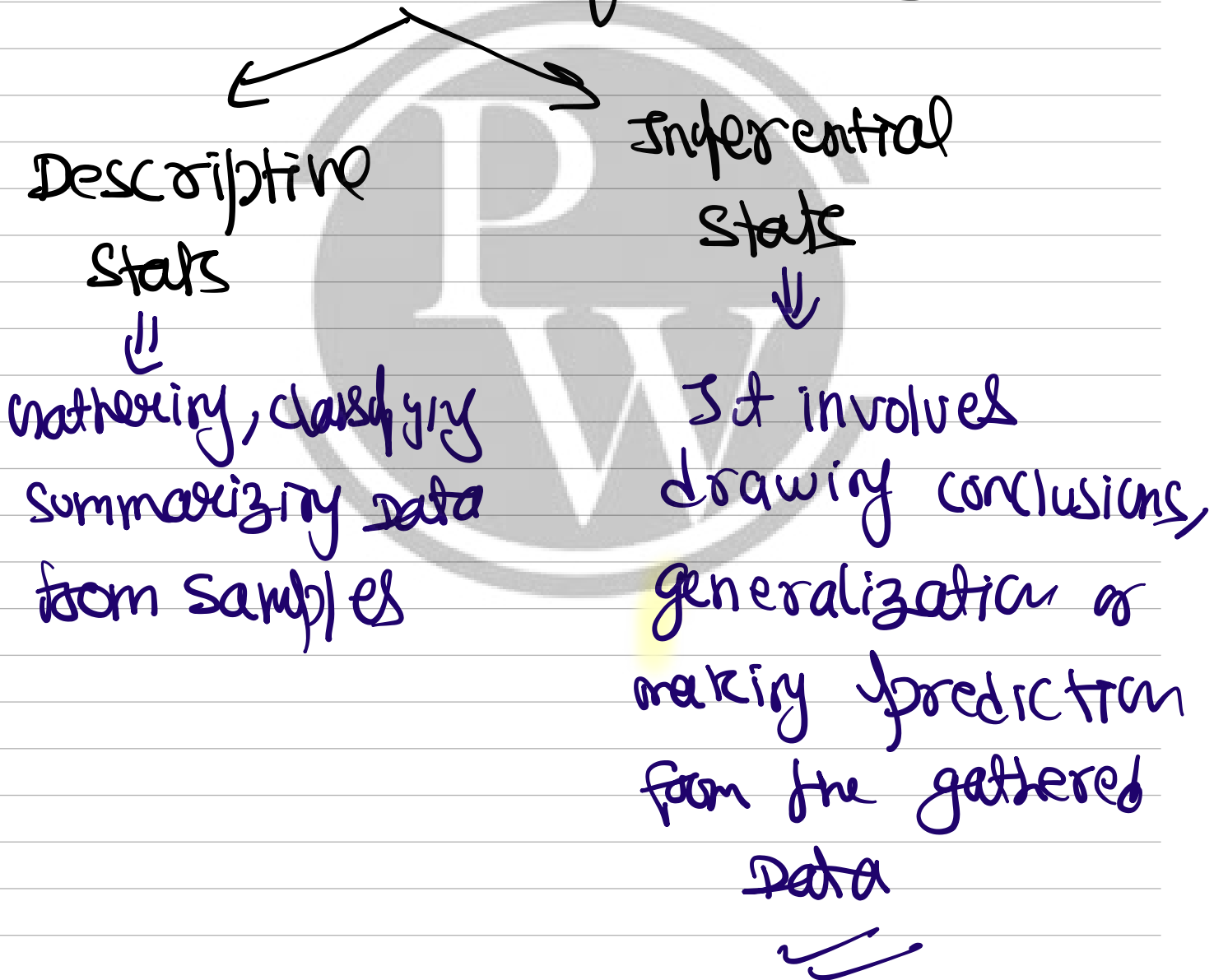


Sampling

By Anurag Chauhan

Branches of Statistics



Population

→ Group of people

→ Group of objects

→ Group of events & observations.

g Height of male students

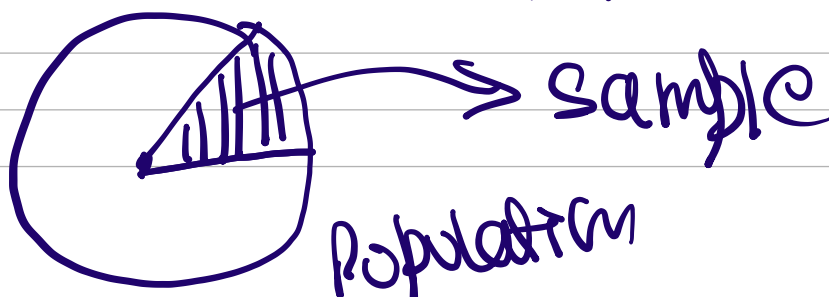
g Blood pressure of females
b/w age 40 & 60 years

g 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

g Population mean = μ
Population variance = σ^2

Statistics : measurable characters of sample

g 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

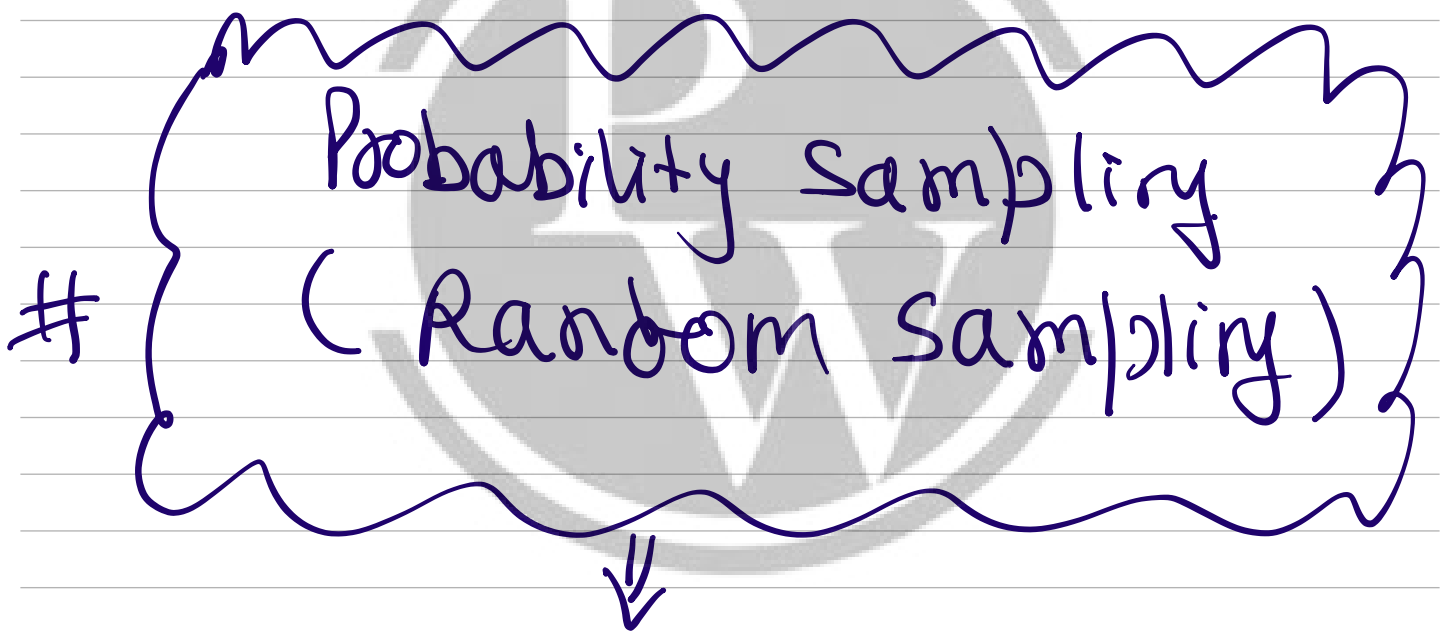
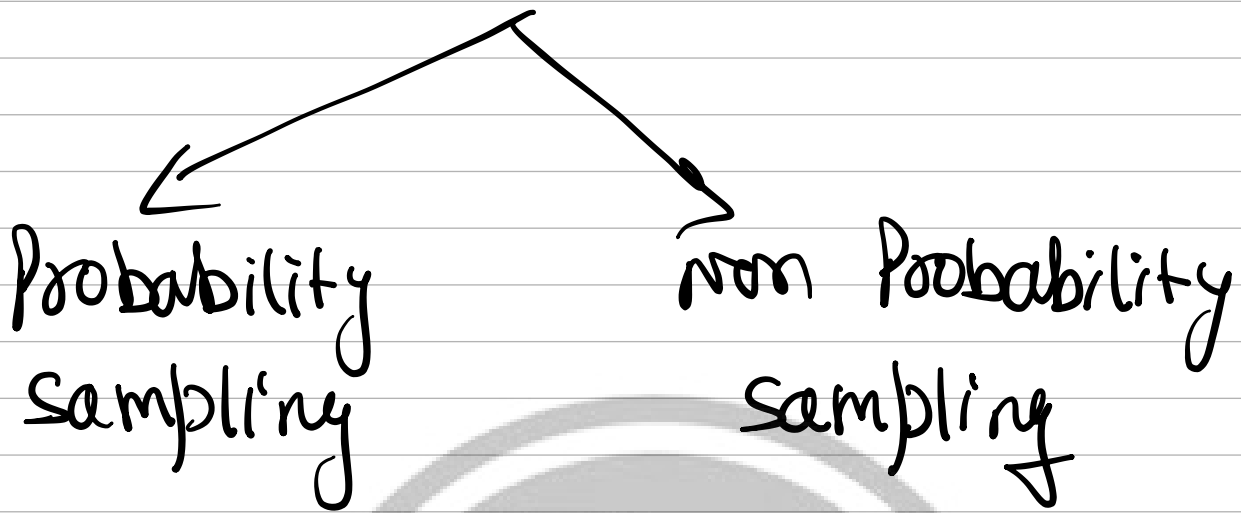
1) **Speed** : Sampling takes less time

2) **Cost** : Sampling is least expensive

3) **Accuracy & Reliability** : Census gives more Accurate & Reliable results

unless there is no bias or error in collecting information.

Types of sampling



When elements are
Randomly selected
Fair 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.

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

ii) (Quasi Random sampling) Systematic Sampling

In this sampling every member of the population is assigned a number. This first member

is selected randomly & then instead of choosing other randomly we choose them in regular intervals

ii) There are 1000 employees in a company, they are assigned numbers from 1 to 1000. We randomly select number '6' from first 10 numbers & after that every 10th 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 subgroup.

g: 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) cluster sampling (multi stage sampling)

When population size is large, divide the population in subgroups (each subgroup has similar characteristics 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 work employees of Nagar Nigam.

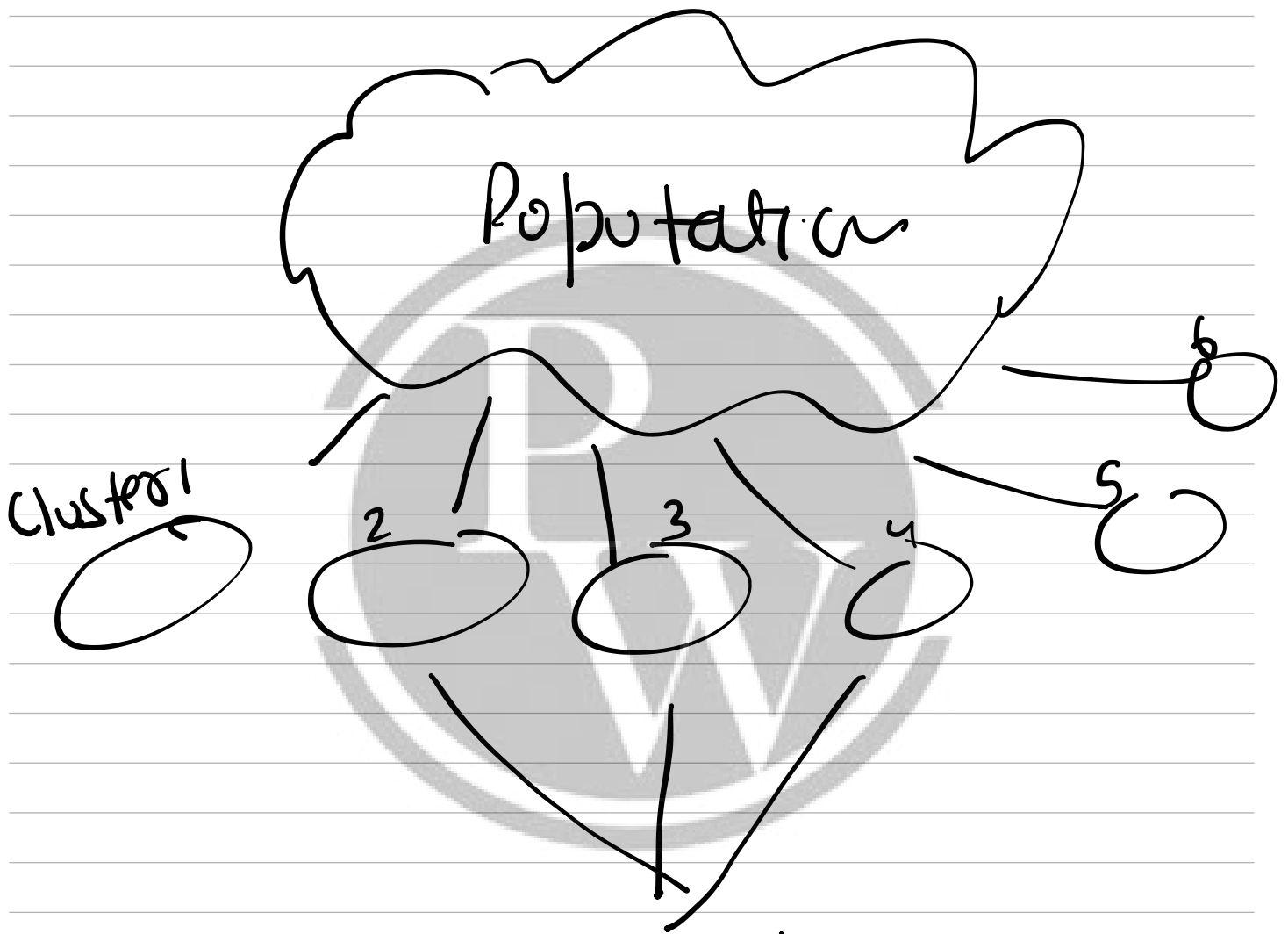
Nagar Nigam Dept.

of each state is a cluster. (sample)

Then you select any 3 departments 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

of Indian idol

2) Convenience sampling

those elements are selected which are easily accessible to researcher.

g) 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) Snowball 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

→ faulty 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 bits

→ Wrong measurements

→ Untrained interviewer

→ Biased opinion

Population

Aggregation of all units
under consideration

Population size
denoted by 'N'

Finite
Population

finite countable
elements

Infinite
Population

Uncountable
elements

Population

Existent

- Population of a town
- Car produced by Hyundai

Hypothetical
(Imaginary)

- A coin is tossed 20 times
- 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 (μ) =
$$\frac{\sum_{i=1}^N x_i}{N}$$

⇒ Population Proportion (P) =
$$\frac{x}{N}$$

g In a population 10,000 units
1000 units are defective

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

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

Statistics

measurable characters of sample

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

$$\hat{\sigma}_x = \text{Sample S.D} = s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}$$

$$\hat{p} = \text{Sample proportion} = p = \frac{x}{n}$$

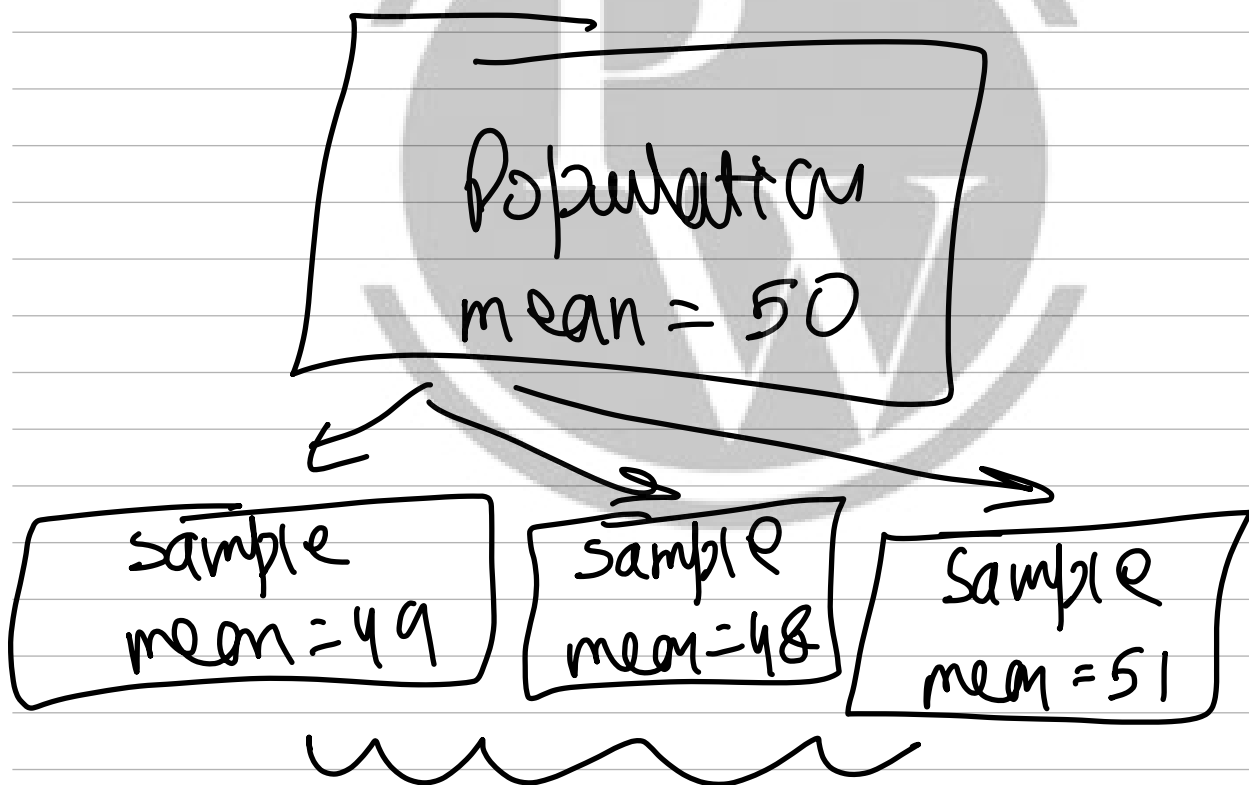
Total no of samples
with replacement $= (N)^n$

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, similarly SD, correlation also can be different.

This is sampling fluctuations.

g 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

$$\text{Total samples} = {}^N C_n = {}^4 C_2 = \frac{4!}{2!2!} = 6$$

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

$$\text{Total samples} = n^n = 4^2 = 16$$

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 वस याद रख लेना

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

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

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

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

SRSWA → Simple Random Sampling with Replacement

SRSWOR → Simple Random Sampling without Replacement

Standard Error for Proportion

$$SE(P) = \sqrt{\frac{Pq}{n}}$$

SRSWR

$$SE(P) = \sqrt{\frac{Pq}{n}} \sqrt{\frac{N-n}{N-1}}$$

SRSWOR

$\sqrt{\frac{N-n}{N-1}}$ = 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

$$N=3, \quad n=2$$

$$\text{Total samples} = N^n = 3^2 = 9$$

<u>sample</u>	<u>Sample mean (\bar{x})</u>
1, 1	1
1, 5	3
1, 3	2
5, 1	3
5, 5	5
5, 3	4
3, 1	2
3, 5	4

$$3, 3 \mid 3$$

For sampling Distribution
sample mean is considered
Random variable

X: Sample mean = 1, 2, 3, 4, 5

x_i	P_i
1	$\frac{2}{9}$
2	$\frac{2}{9}$
3	$\frac{3}{9}$
4	$\frac{2}{9}$
5	$\frac{1}{9}$
	1

Sampling Distribution

ii) > without Replacement

Population { 1, 5, 3 }

Sample of two elements

$$\text{Total no of sample} = {}^N C_n = {}^3 C_2 = 3$$

Sample	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	$\frac{1}{3}$
3	$\frac{1}{3}$
4	$\frac{1}{3}$
	1

Sampling Distribution

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

variance of Sampling Distribution = $\sum p_i x_i^2 - (\sum p_i x_i)^2$

Q

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
		2.1	4.9

$$\text{mean or Expectation} = E(x) = \sum p_i x_i = 2.1$$

$$\begin{aligned} \text{variance} &= \sum p_i x_i^2 - (\sum p_i x_i)^2 \\ &= 4.9 - (2.1)^2 \\ &= 4.9 - 4.41 \\ &= 0.49 \end{aligned}$$

$$\begin{aligned} \text{S.D or S.E} &= \sqrt{0.49} \\ &= 0.7 \end{aligned}$$

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) without replacement

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

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

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

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