SAMPLING METHOD

IIBF ABM

SAMPLING METHOD

- Decision making in a situation involves collecting information and then using this data for the future strategy.
- Usually, we cannot use the complete data because of the sheer size or numbers involved.
- Therefore, we take a sample and test it.
- The word sample is used to describe a portion chosen from the population.
- Statisticians use the word population to refer not only to people but to all items that are to be studied.

For example, if a milk plant processes 1 lakh litres of milk every day, one cannot break open each packet and test the milk for quality. Here we take samples from each batch.

Important points

- We can describe samples and populations by using measures such as mean, median, mode and standard deviation.
- When these terms describe a sample, they are called statistic and are not from the population but estimated from the sample.
- When these terms describe a population, they are called parameters.
- A statistic is a characteristic of a sample; a parameter is a characteristic of a population,
- statisticians use lower case Roman letters to denote sample statistics and Greek or Capital letters to denote population parameters.

	Population	**Sample**
Definition	**Collection of all items**	**Part of the population**
Characteristics	**Parameters**	**Statistics**
Symbols	**Size - N**	**Size – n**
	Mean - μ	**Mean - $\bar{x}$**
	Standard Deviation - σ	**Standard Deviation - s**

TYPES OF SAMPLING

1. Judgement or Non-random Sampling
2. Random Sampling

Judgement or Non-random Sampling

- In judgement sampling, personal knowledge or opinions are used to identify the items from the population that are to be included in the sample.
- A sample selected by judgement sampling is based on someone's experience with the population.
- An oil drilling company would ask an experienced geologist to test different terrains or land beneath the sea before deciding where to explore for oil.

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

Random Sampling

In probability sampling all the items in the population have a chance of being chosen in the sample.

I. Simple Random Sampling
II. Systematic random sampling
III. Stratified sampling
IV. Cluster sampling

1. Simple Random Sampling

Simple Random Sampling selects samples by methods that allow each possible sample to have an equal probability of being picked and each item in the entire population to have an equal chance of being included in the sample.

Suppose we have 4 teenagers participating in a talk show. We want a sample of two teenagers at a time for participating with the chat show host.

Teenagers A, B, C, D

Possible samples of two teenagers: AB, AC, AD, BC, BD, CD

Probability of drawing this samples of two people is the same as below:

P(AB) = 1/6

P(AC) = 1/6

P(AD) = 1/6

P(BC) = 1/6

P(BD) = 1/6

P(CD) = 1/6

one student appears in 3 of the 6 possible samples. Therefore, probability of a particular student in the sample is

P(A) = 1/2

P(B) = 1/2

P(C) = 1/2

P(D) = ½

2. Systematic Sampling

In systematic sampling, elements are selected from the population at a uniform level that is measured in time, order, or space.

3. Stratified Sampling

- To use stratified sampling, we divide the population into relatively homogenous groups, called strata.
- Then we use one of two approaches.
- Either we select at random from each stratum a specified number of elements corresponding to the proportion of that stratum in the population as a whole
- or we draw an equal number of elements from each stratum and give weight to the results according to the stratum; proportion of total population.
- With either approach, stratified sampling guarantees that every element in the population has a chance of being selected

When to use Stratified Sampling

- Stratified sampling is appropriate when the population is already divided into groups of different sizes.
- Example - middle class, upper class, lower middle class, etc. or according to age, race, sex or any other stratification.
- A jeans company may want to study which age group prefers to wear jeans the maximum. Thus, the age groups may be 13 to 20, 20 to 30, 30 to 40.

4. Cluster Sampling

In cluster sampling, we divide the population into groups or clusters and then select a random sample of these clusters.

If a market Research team is attempting to determine by sampling the average number of television sets per household in a large city, they could use a city map and divide the territory into blocks and then choose a certain number of blocks (clusters) for interviewing.

Comparison of Stratified and Cluster Sampling

- With both stratified and cluster sampling, the population is divided into well-defined groups.
- We use stratified sampling when each group has small variation within itself but there is wide variation between the groups.
- We use cluster sampling in the opposite case - when there is considerable variation within each group but the groups are essentially similar to each other.

SAMPLING DISTRIBUTION

- Each sample you draw from a population would have its own mean or measure of central tendency, and standard deviation.
- Thus, the statistics we compute for each sample would vary and be different for each random sample taken.

Describing Sampling Distributions

For example:

We take a finite population of 5 young boys; A, B, C, D, E and collect data about their heights in centimeters.

Boy	*A*	*B*	*C*	*D*	*E*
HEIGHT (cm)	*160*	*162*	*164*	*170*	*156*

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

Now, if we take samples of size 3 that is, select 3 boys in each sample, we will get 10 different samples.

No	1	2	3	4	5	6	7	8	9	10
Sample	ABC	ABD	ABE	BCD	BCE	ACD	ACE	ADE	BDE	CDE
DATA	160, 162, 164	160, 162, 170	160, 162, 156	162, 164, 170	162, 164, 156	160, 164, 170	160, 164, 156	160, 170, 156	162, 170, 156	164, 170, 156
Mean	162	164	159.33	165.33	160.66	164.66	160	162	162.66	163.33

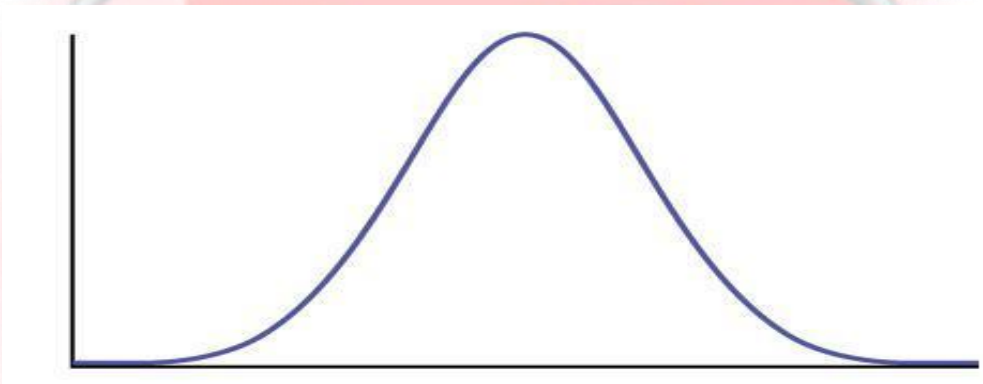
Concept of Standard Error

- Standard deviation of the distribution of the sample means is called the standard error of the mean,
- Similarly standard error of the proportion is the standard deviation of the distribution of the sample proportions.
- The standard deviation of the sampling distribution of means measures the extent to which the means vary because of a chance error in the sampling process.

- Thus, the standard deviation of the distribution of a sample statistic is known as the standard error of the statistic.

NORMAL DISTRIBUTION

Normal distribution, also known as the Gaussian distribution, is **a probability distribution that is symmetric about the mean**



In that case:

$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$

$\mu = \bar{x}$

	Property	Equation
Properties of Sampling Distribution of Mean, when population is normally distributed.	• **Sampling distribution has mean equal to population mean.** • **Sampling distribution has standard deviation equal to**	$\mu_{\bar{x}} = \bar{X}$ $\sigma_{\bar{x}} = \sigma/\sqrt{n}$

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

	population standard deviation divided by square root of sample size.	

SAMPLING FROM NON-NOMRAL POPULATIONS

- when the population is normally distributed, the sampling distribution of the mean is also normal.
- But many populations that are not normally distributed.
- How does the sampling distribution of the mean behave when the population from which the samples are drawn is not normal?

Concerning five motorcycle owners and the lives of their tyres. Because only five people are involved, the population is too small to be approximated by a normal distribution.

Owner	C	D	E	F	G
Tyre Life in months	3	3	6	9	15

Total life = 36 months

Mean = 36/5 = 7.2 months

Samples of Three	Sample Data	Sample Mean
EFG	6 + 9 + 15	10
DFG	3 + 9 + 15	9
DEG	3 + 6 + 15	8
DEF	3 + 6 + 9	6
CFG	6 + 6 + 9	7
CEG	3 + 6 + 15	8
CEF	3 + 6 + 9	6
CDF	3 + 3 + 9	5
CDE	3 + 3 + 9	5
CDG	3 + 3 + 15	8

Total = 72 months

Mean = 7.2 months

CENTRAL LIMIT THEOREM

- mean of the sampling distribution is equal with the population mean
- As sample size increases the, the sampling distribution will approach normality regardless the shape of population distribution
- **Relationship between the shape of the population distribution and shape of the sampling distribution of the mean is called the Central Limit of Distribution**

Significance: It permits us to use the sample statistics to make inferences about the **population parameters** without knowing the shape of the frequency distribution of the population

- Distribution of the sample means tends to a normal distribution regardless the shape of the population distribution.

RELATIONSHIP BETWEEN SAMPLE SIZE AND STANDARD ERROR

- If the dispersion of the sample means around the population mean decreases, it means the values taken by sample mean tends to approach the population mean
- As Standard Error decreases means the value of any sample mean will probably be closer to the population mean.

FINITE POPULATION MULTIPLIER

In case of finite population without replacement formula used for calculations is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{(N-n)}{N-1}}$$

QUESTION AND ANSWER

Q1. In a sample of 25 observations from a normal distribution with mean 98.6 and standard deviation 17.2. What is Probability that the sample mean will lie between 92 and 102?

Ans: 0.8115

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.01736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3565	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.03997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4457	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4938	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4915
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Q2. A bank calculates that its individual savings accounts are normally distributed with a mean of Rs.2000 and a standard deviation Rs.600. If the bank takes a random sample of 100 accounts, what is the probability that the sample mean will lie between & 1,900 and Rs. 2,050?

Ans: 0.7482

Q3. The data relating to sample distribution of annual earnings of bank employees with five years' experience is given. This distribution has a mean of Rs 19,000 and a standard deviation of Rs 2,000. If we draw a random sample of 30 employees, what is the probability that their earnings will average more than Rs 19,750 annually?

Ans: 0.0200 or 2%

Q4.A portion of the elements in a population chosen for direct examination or measurement is known as ________.

Ans: Sample

Mail: banishagupta2099@gmail.com
Website: https://iibf.info

Q5. Under ____________ inferences about a population are made from information within itself but there is wide variation between different groups.

Ans: Stratified method

Q6. A method of random sampling in which elements are selected from the population at uniform intervals is called _________

Ans: systematic sampling

Q7. Within a population, groups that are like each other are called _________?

Ans: clusters

Q8. Choose the pair of symbols that is not describing ___________ as a parameter and ___________ as a statistic.

(a) s, μ

(b) σ, s

(c) N, n

(d) All of these

Ans: s, μ

Q9. Suppose that a population with N = 144 has μ = 24. What is the mean of the sampling distribution of the mean for samples of size 25?

Ans: 24

Q10. Large samples are always a good idea because they decrease the standard error. (True/False)

Ans: True