Unit-1 :-

Inferential Statistics

Page No.
Date.

Que- Define Data analysis and Data analytics with Difference between both?

⇒

Data analysis :-

Data analysis is the process of inspecting, cleaning, transforming and modeling data with the goal of discovering useful information, informing conclusions and supporting decision-making.

Data analytics :-

- Data analytics is the process of analyzing raw data to find trends and answer questions.

	Data Analytics	Data Analysis
1	It is described as a traditional form or generic form of analytics.	It is described as a particularized form of analytics.
2	It includes several stage like the collection of data and then the inspection of business data is done.	To process data, firstly raw data is defined in a meaningful manner, then data cleaning and conversion are done to get meaningful information from raw data.
3	It supports decision making by analyzing enterprise data	It analyzes the data by focusing on insights into business data.

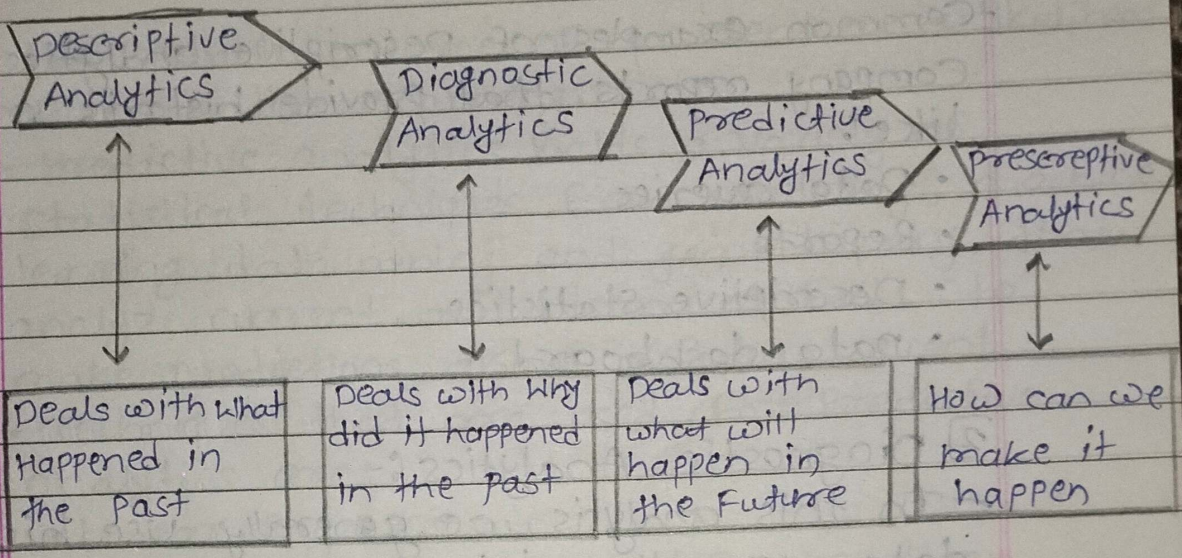
4] It uses various tools to process data such as Tableau, python, Excel etc.	4] It uses different tools to analyze data such as Rapid miner, open Refine, Node XL, KNIME etc.
5] Descriptive analysis cannot be performed on this	5] A descriptive analysis can be performed on this.
6] one can find anonymous relations with the help of this	6] one cannot find anonymous relations with the help of this
7] It does not deal with inferential analysis	7] It supports inferential analysis.

Que- Explain Types of analytic with example.

⇒

Types of Analytics:-

1] Descriptive Analytics
2] Diagnostic analytics
3] Predictive analytics
4] Prescreptive analytics



1] Descriptive Analytics (business intelligence and data mining)

- Descriptive analytics looks at data and analyze past event for insight as to how to approach future events.
- It looks at past performance and understands the performance by mining historical data to understand the cause of success or failure in the past.
- Almost all management reporting such as sales, marketing, operations and finance uses this type of analysis.
- The descriptive model quantifies relationships in data in a way that is often used to classify customers or prospects into groups.
- Unlike a predictive model that focuses on predicting the behavior of a single customer, descriptive analytics identifies many different relationships between customer and product.

- Common examples of Descriptive analytics are Company reports that provide historic reviews like:
  - Data queries
  - Reports
  - Descriptive statistics
  - Data dashboard

2] Diagnostic Analytics :-
- In this analysis, we generally use historical data over other data to answer any question or for the solution of any problem.
- We try to find any dependency and Pattern in the historical data of the Particular Problem.
- For example, companies go for this analysis because it gives a great insight into a Problem, and they also keep detailed information about their disposal otherwise data collection may turn out individual for every Problem and it will be very time-consuming.
- common techniques used for Diagnostic Analytics are:
  - Data discovery
  - Data mining
  - correlations

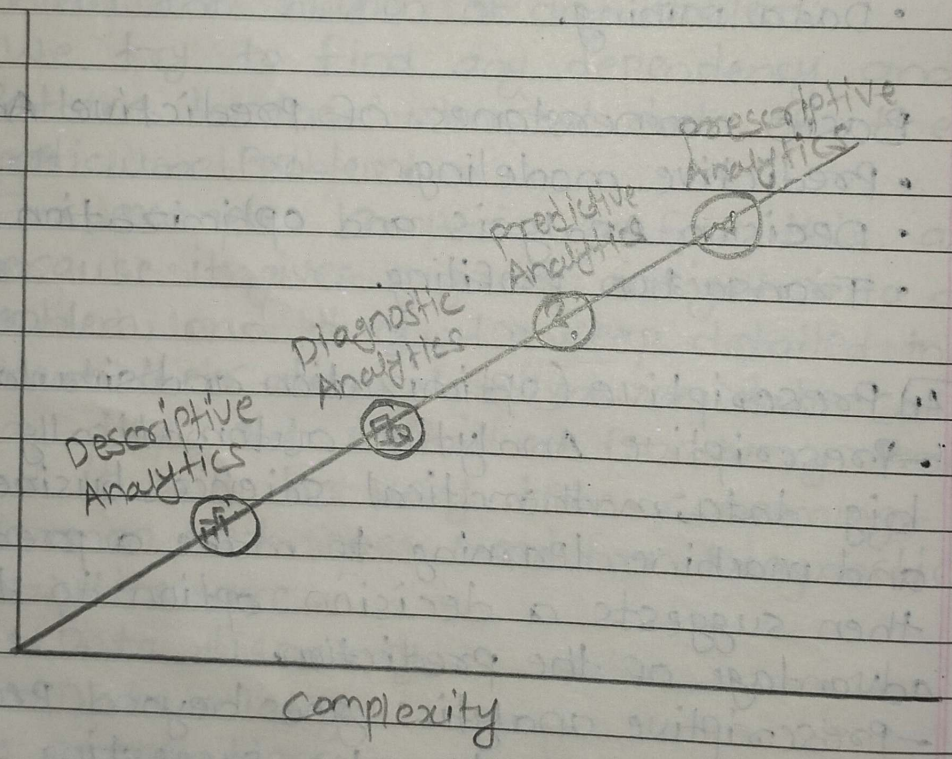
3] Predictive (Forecasting) :-
- Predictive analytics turn the data into valuable, actionable information.
- Predictive analytics uses data to determine

the probable outcome of an event or a likelihood of a situation occurring.
- Predictive analytics holds a variety of statistical techniques from modeling, machine learning, data mining and game theory, that analyze current and historical facts to make predictions about a future event.
- Techniques that are used for Predictive analytics are:
  - Linear Regression
  - Time series Analysis and Forecasting
  - Data mining.

Basic cornerstones of Predictive Analytics
- Predictive modeling
- Decision Analysis and optimization
- Transaction profiling.

4) Prescriptive (optimization and simulation) :-
- Prescriptive Analytics automatically synthesize big data, mathematical science, business rule and machine learning to make a prediction and then suggests a decision option to take advantage of the prediction.
- Prescriptive analytics goes beyond Predicting Future outcomes by also suggesting action benefits from the predictions and showing the decision maker the implication of each decision option.
- prescriptive Analytics not only anticipates what will happen and when to happen but

also why it will happen.
- Further, prescriptive Analytics can suggest decision options on how to take advantage of a future opportunity or mitigate a future risk and illustrate the implication of each decision option.
- For example, Prescriptive Analytics can benefit healthcare strategic planning by using analytic to leverage operational and usage data combined with data of external factors such as economic data, population demography etc.



Page No.
Date

Que- Discuss the need of data analytics?

⇒

need of data analytics :-
Data analytics is used for several reasons :-

1) Informed Decision making :-
- Data analytics enables organizations to make informed decisions based on insights derived from data.
- By analyzing patterns and trends, businesses can identify opportunities, optimize processes, and mitigate risks.

2) Competitive Advantage :-
- In today's competitive landscape, organizations need to stay ahead of the curve.
- Data analytics allows businesses to gain a competitive edge by understanding customer behavior, market trends and industry dynamics better than their competitors.

3) Improved Efficiency :-
- By analyzing operational data, organizations can identify inefficiencies and bottlenecks in their processes.
- This allows them to streamline operations, reduce costs, and improve overall efficiency.

4) Personalized customer Experiences :-
- Data analytics helps businesses understand their customers on a deeper level.

- By analyzing customer data, organizations can personalize marketing campaigns, tailor Products and services to specific needs and enhance customer Satisfaction and loyalty.

5) Risk management :-
- Data analytics can help organizations identify and mitigate risks effectively.
- By analyzing historical data and using predictive analytics, businesses can anticipate potential risks and take proactive measures to mitigate them.

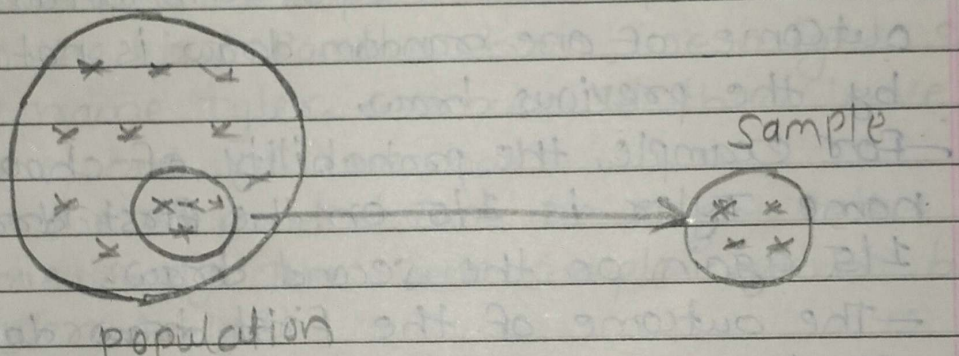
6) Innovation :-
- Data analytics fuels innovation by uncovering insights that lead to new Products, services and business models.
- By analyzing data from various sources, organizations can identify emerging trends and opportunities for innovation.

Page No.
Date.

Que- What is meant by sample? Explain sampling with replacement and without replacement with proper example.

→

Sample :-
- A sample is an unbiased number of observations taken from a population.
- Sample implies a part or unit taken at random from a larger whole and so presumed to be typical of its qualities.



- Sampling With Replacement :-
- suppose we have the names of 5 students in a hat :
  - Andy
  - karl
  - Tyler
  - Becca
  - Jessica

- suppose we would like to take a sample of 2 students with replacement.

- on the first random draw, we might select the name Tyler. we would then place his name back in the hat and draw again.
- on the second draw, we might select the name Tyler again.
- Thus our sample would be: {Tyler, Tyler}
- This is an example of obtaining a sample with replacement because we replace the name we choose after each random draw.
- When we sample with replacement, the items in the sample are independent because the outcome of one random draw is not affected by the previous draw.
- for example, the probability of choosing the name Tyler is 1/5 on the first draw and 1/5 again on the second draw.
- The outcome of the first draw does not affect the probability of the outcome on the second draw.
- sampling with replacement is used in many different scenarios in statistics and machine learning, including:
  - Bootstrapping
  - Bagging
  - A simple Introduction to Boosting in machine learning
  - A simple Introduction to Random Forests.

- In each of these methods, sampling with replacement is used because it allows us to use the same dataset multiple times to build models as opposed to going out and gathering new data, which can be time-consuming and expensive.

- Sampling without Replacement :-
- suppose we have the names of 5 students in a hat :-
  - Andy
  - Karl
  - Tyler
  - Becca
  - Jessica

- suppose we would like to take a sample of 2 students without replacement.
- on the first Random draw, we might select the name Tyler. we would then leave his name out of the hat.
- on the second draw, we might select the name Andy. Thus our sample would be: {Tyler, Andy}

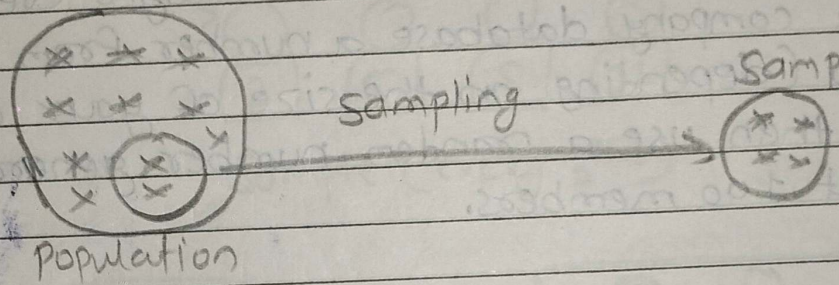
- This is an example of obtaining a sample without replacement because we do not replace the name we choose after each random draw.
- When we sample without replacement, the items in the sample are dependent because the outcome of one random draw is affected by the previous draw.
- For example, the probability of choosing the name Tyler is 1/5 on the first draw and the probability of choosing the name Andy is 1/4 on the second draw.

- The outcome of the first draw affects the probability of the outcome on the second draw.
- Sampling without replacement is the method we use when we want to select a random sample from a population.
- For example, if we want to estimate the median household income in cincinnati, ohio there might be a total of 500,000 different households.
- Thus, we might want to collect a random sample of 2,000 households but we don't want the data for any given household to appear twice in the sample so we would sample without replacement.
- In other words, once we've chosen a certain household to be included in the sample we don't want there to be any chance of selecting that household to be included again.

Que- Classify sampling types. Briefly explain probability sampling Techniques with proper diagram.

⇒

Sampling :-
- It is the practice of selecting an individual group from a population to study the whole population.

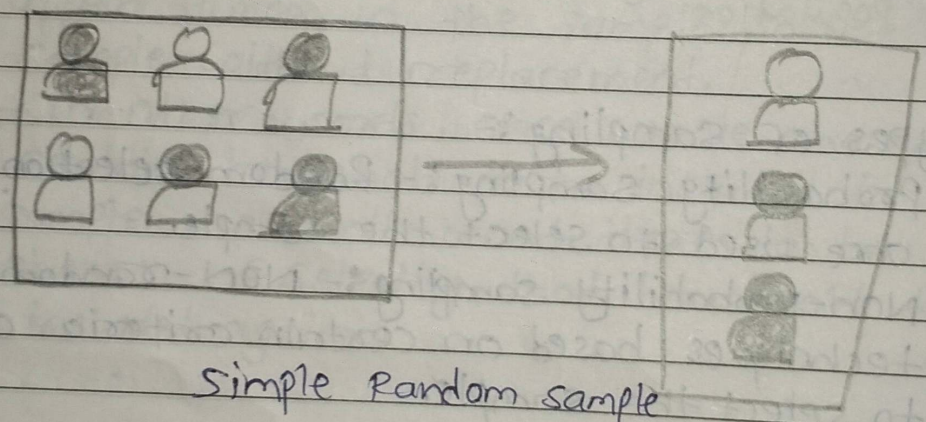


Types of sampling :-
1) Probability Sampling :- Random selection techniques are used to select the sample.
2) Non-probability sampling :- NON-random selection techniques based on certain criteria are used to select the sample.

1) probability sampling Techniques :-
- probability sampling techniques are one of the important types of sampling techniques.
- probability sampling allows every member of the population a chance to get selected.
- It is mainly used in quantitative research when you want to produce results representative of the whole population.

a) Simple Random Sampling :-

- In simple random sampling, the researcher selects the participants randomly.
- There are a number of data analytics tools like random number generators and random number tables used that are based entirely on chance.
- Example :- The researcher assigns every member in a company database a number from 1 to 1000 (depending on the size of your company) and then use a random number generator to select 100 members.

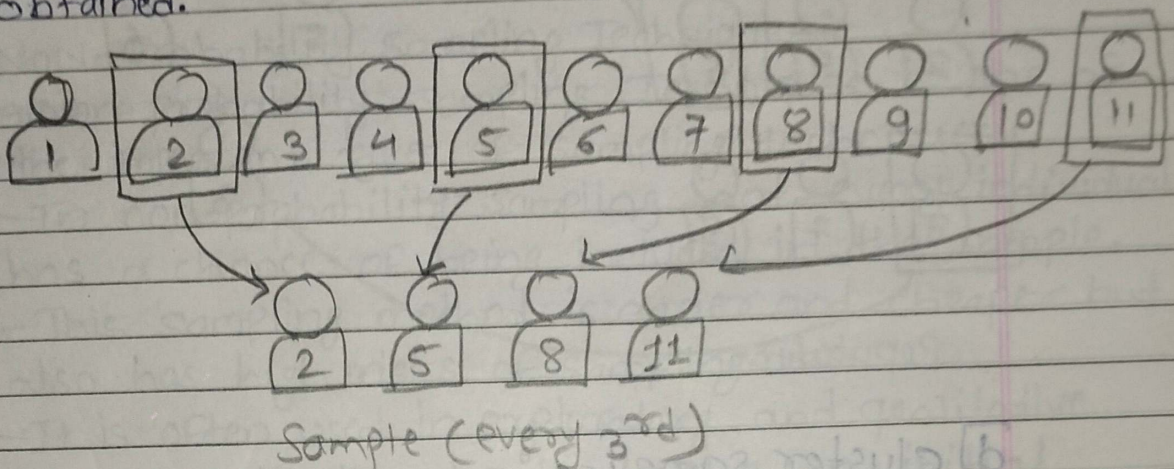


Simple Random sample

b) Systematic Sampling :-

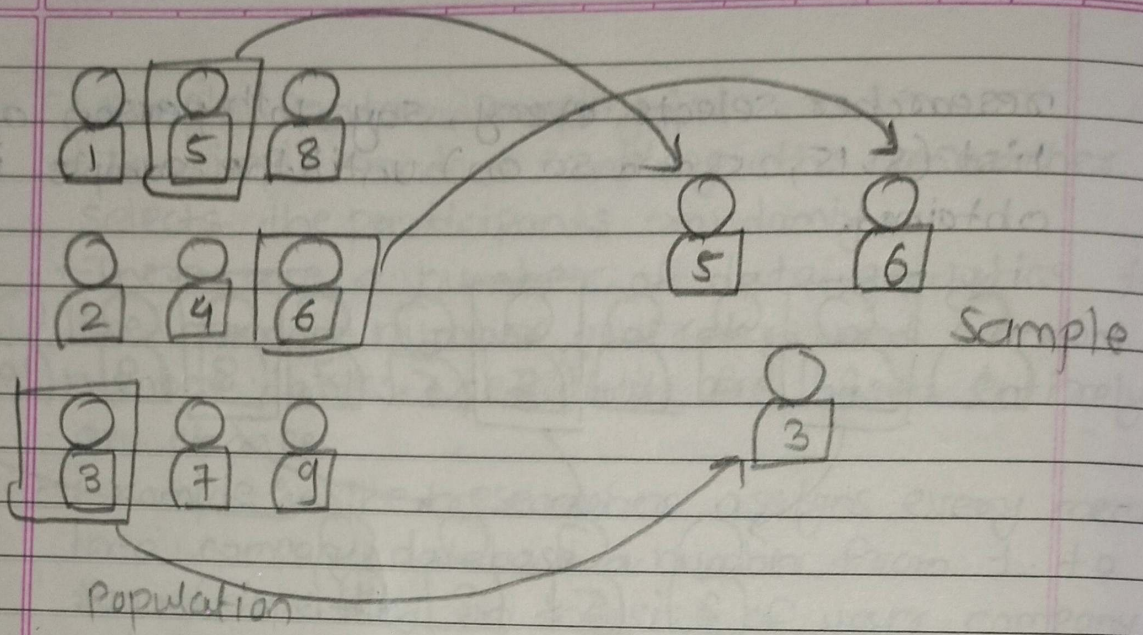
- In systematic sampling, every population is given a number as well like in simple random sampling.
- However, instead of randomly generating numbers, the samples are chosen at regular intervals.
- Example :-
- The researcher assigns every member in the company database a number. Instead of randomly generating numbers, a random starting point (say 5) is selected. From that number onwards, the

researcher selects every, say, 10th person on the list (5, 15, 25 and so on) until the sample is obtained.



c) stratified sampling :-

- In stratified sampling, the population is sub-divided into subgroups, called strata, based on some characteristics (age, gender, income).
- After forming a subgroup, we can then use random or systematic sampling to select a sample for each subgroup. This method allow you to draw more precise conclusions because it ensures that every subgroup is properly represented.
- Example :-
- If a comany has 500 male employees and 100 female employees, the researcher wants to ensure that the sample reflects the gender as well. so the population is divided into two subgroups based on gender.

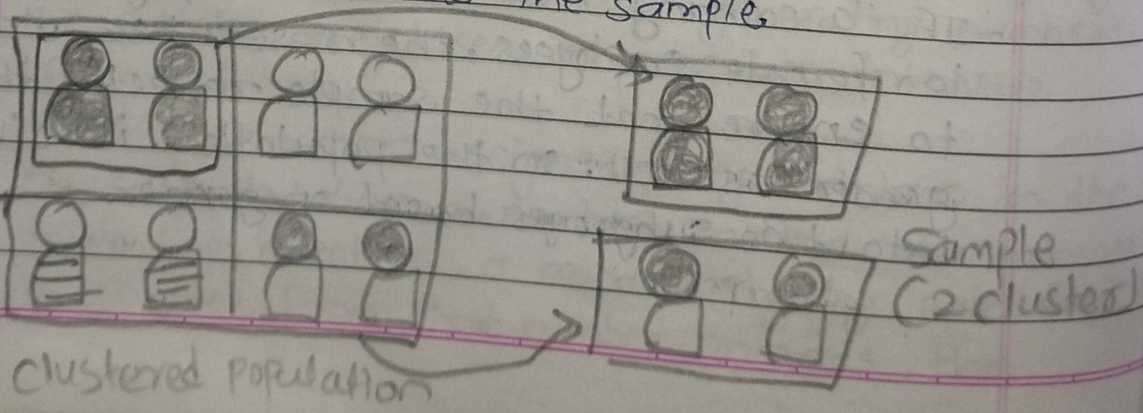


d) cluster sampling :-

- In cluster sampling, the population is divided into subgroups, but each subgroup has similar characteristics to the whole sample.
- Instead of selecting a sample from each subgroup, we randomly select an entire subgroup.
- This method is helpful when dealing with large and diverse populations.

- Example:-

A company has over a hundred offices in ~~ten~~ cities across the world which has roughly the same number of employees in similar job roles. The reasearcher randomly selects 2 to 3 offices and uses them as the sample.



Que- Explain non-probability sampling Techniques with proper diagram.

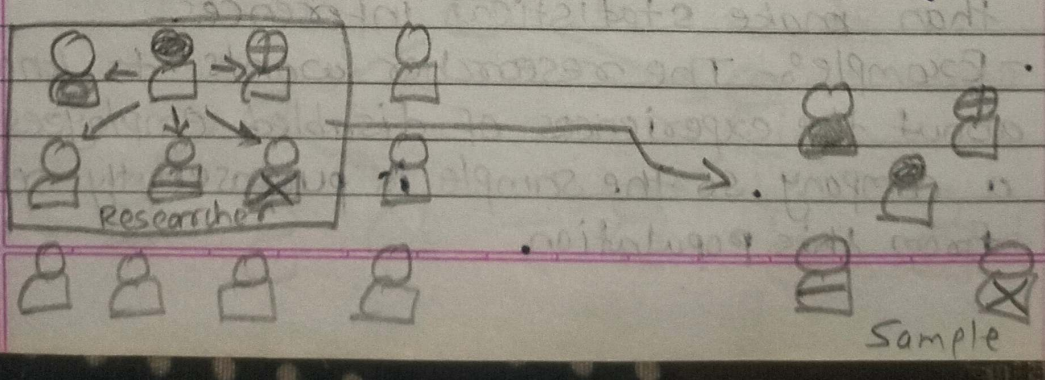
⇒

NON-Probability Sampling Techniques :-

- Non-probability sampling Techniques is one of the important types of sampling techniques.
- In non-probability sampling not every individual has a chance of being included in the sample.
- This sampling method is easier and cheaper but also has high risks of sampling bias.
- It is often used in exploratory and qualitative research with the aim to develop an initial understanding of the population.

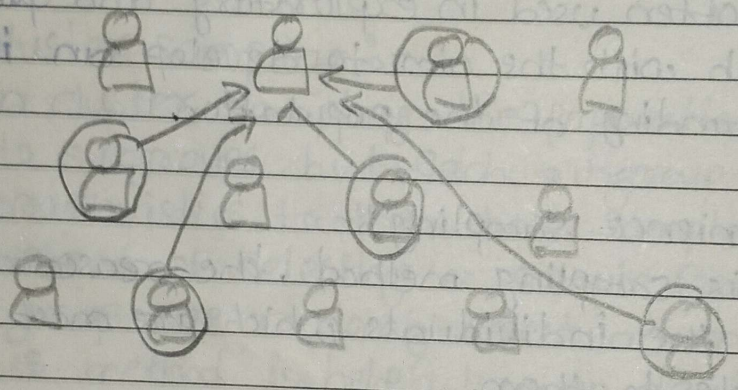
a) Convenience Sampling :-

- In this sampling method, the reasearcher simply selects the individuals which are most easily accessible to them.
- This is an easy way to gather data, but there is no way to tell if the sample is representative of the entire population.
- The only criteria involved is that people are available and willing to participate.
- Example :- The researcher stands outside a company and asks the employees coming in to answer question or complete a survey.

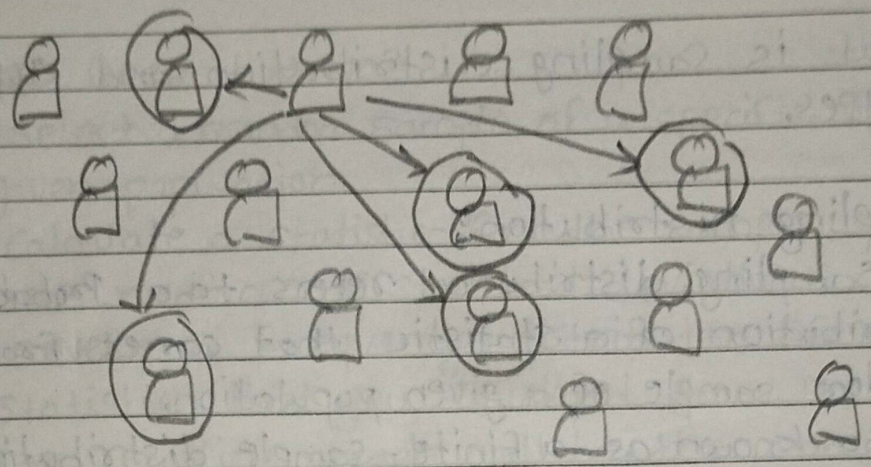


Page No.
Date.

2] Voluntary Response sampling :-
- Voluntary response sampling is similar to convenience sampling, in the sense that the only criterion is people are willing to participate.
- However, instead of the researcher choosing the participants, the participants volunteer themselves.
- Example :-
The researcher sends out a survey to every employee in a company and gives them the option to take part in it.

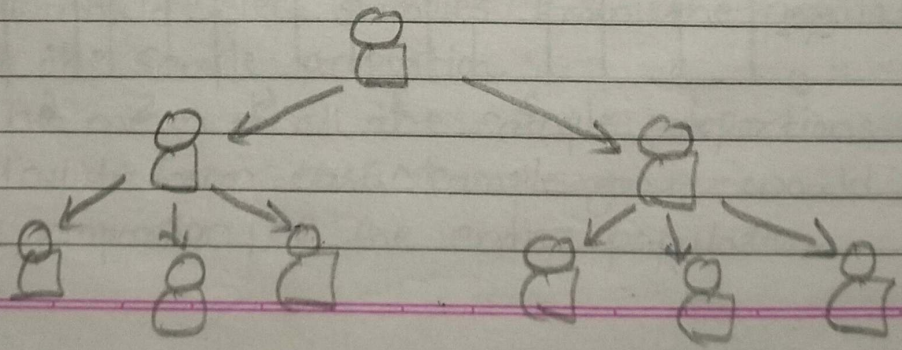


3] Purposive sampling :-
- In purposive sampling, the researcher uses their expertise and judgment to select a sample that they think is the best fit.
- It is often used when the population is very small and researcher only wants to gain knowledge about a specific phenomenon rather than make statistical inferences.
- Example :- The researcher wants to know about the experiences of disabled employees at a company. So the sample is purposefully selected from this population.



4] Snowball Sampling :-

- In snowball Sampling, the research participants recruit other participants for the study.
- It is used when participants required for the research are hard to find.
- It is called snowball sampling because like a snowball, it picks up more participants along the way and gets larger and larger.
- Example :- The reasearcher wants to know about the experiences of homeless people in a city. since there is no detailed list of homeless people, a probability sample is not possible.
- They only way to get the sample is to get in touch with one homeless person who will then put you in touch with other homeless people in a particular area.



Page No.
Date.

Que- What is Sampling distribution and explain its types.

⟹

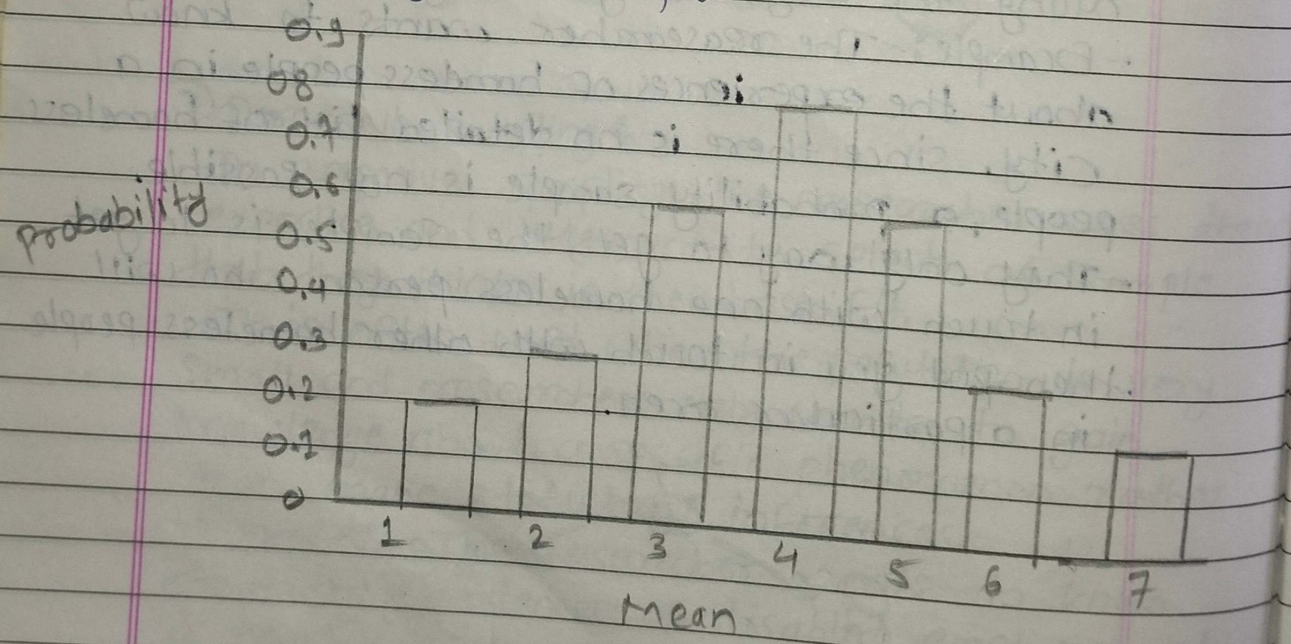
Sampling Distribution:-

-A Sampling distribution refers to a Probability distribution of a statistic that comes from choosing random sample of a given population.

-Also known as a finite-sample distribution, it represents the distribution of frequencies on how spread apart various outcomes will be for a specific population.

-The sampling distribution depends on multiple factors - the statistic, sample size, sampling process and the overall population.

-It is used to help calculate statistics such as means, ranges, variances and standard deviations for the given sample.



How Does it Work?

1] Select a random sample of a specific size from a given population.

2] Calculate a statistic for the sample, such as the mean, median or standard deviation.

3] Develop a frequency distribution of each sample statistic that you developed from the step above. The resulting graph will be the sampling distribution.

- Types of sampling distribution:-

1] Sampling distribution of mean:-

- As shown from the example above, you can calculate, the mean of every sample group chosen from the population and plot out all the data points.
- The graph will show a normal distribution, and the center will be the mean of the sampling distribution, which is the mean of the entire population.

2] Sampling distribution of proportion:-

- It gives you information about proportions in a population.
- You would select samples from the population and get the sample proportion.
- The mean of all the sample proportions that you calculate from each sample group would become the proportion of the entire population.

3] T-distribution :-

- T-distribution is used when the sample size is very small or not much is known about the population.
- It is used to estimate the mean of the population, confidence intervals, statistical differences and linear regression.

- Practical Example :-

- Suppose you want to find the average height of children at the age of 10 from each continent.
- You take random samples of 100 children from each continent and you compute the mean for each sample group.
- For example, in South America, you randomly select data about the heights of 10 year-old children and you calculate the mean for 100 of the children. You also randomly select data from North America and calculate the mean height for one hundred 10-year-old children.
- As you continue to find the average heights for each sample group of children from each continent you can calculate the mean of the sampling distribution by finding the mean of all the average heights of each sample group. Not only can it be computed for the mean, but it can also be calculated for other statistics such as standard deviation and variance.

Page No.
Date.

Que- Explain the term Hypothesis and illustrate its type with proper example.

⇒

• Hypothesis :-

-In statistics, a hypothesis is defined as a formal statement, which gives the explanation about the relationship between the two or more variables of the specified population.

-It helps the researcher to translate the given problem to a clear explanation for the outcome of the study.

-It clearly explains and predicts the expected outcome.

Types of Hypothesis :-

1) Simple Hypothesis :-

-It shows a relationship between one dependent variable and a single independent variable.

-For example :-

If you eat more vegetables, you will lose weight faster. Here, eating more vegetables is an independent variable, while losing weight is the dependent variable.

2) Complex Hypothesis :-

-It shows the relationship between two or more dependent variables and two or more independent variables.

-For example :- Eating more vegetables and fruits leads to weight loss, glowing skin, and reduces the risk of many diseases such as heart disease.

3) Directional Hypothesis :-
- It shows how a researcher is intellectual and committed to a particular outcome.
- The relationship between the variables can also predict its nature.
- For example - children aged four year eating proper food over a five-year period are having higher IQ levels than children not having a proper meal. This shows the effect and direction of the effect.

3) Non-directional Hypothesis :-
- It is used when there is no theory involved.
- It is a statement that a relationship exists between two variables, without predicting the exact nature (direction) of the relationship.

4) Null Hypothesis :-
- It Provides a statement which is contrary to the hypothesis.
- It's a negative statement, and there is no relationship between independent and dependent variables.
- The symbol is denoted by "Ho".

5) Associative and Causal Hypothesis :-
- Associative hypothesis occurs when there is a change in one variable resulting in a change in the other variable.
- whereas, the causal hypothesis proposes a cause and effect interaction between two or more variables.

Sampling with Replacement $= N^n = 4^4 = 16$

Page No.
Date.

Que- A population consists of four member 3, 7, 11, 15 consider all possible sample size two which can be drawn with replacement from population. Find the population mean, population standard deviation, the mean of sampling distribution of mean and standard deviation of sampling distribution of mean.

⇒

Sr.NO	sample values	Total	sample mean
1	3,7	10	10/2 = 5
2	3,11	14	14/2 = 7
3	3,15	18	18/2 = 9
4	3,3	6	6/2 = 3
5	7,7	14	14/2 = 7
6	7,3	10	10/2 = 5
7	7,11	18	18/2 = 9
8	7,15	22	22/2 = 11
9	11,11	22	22/2 = 11
10	11,3	14	14/2 = 7
11	11,7	18	18/2 = 9
12	11,15	26	26/2 = 13
13	15,15	30	30/2 = 15
14	15,3	18	18/2 = 9
15	15,7	22	22/2 = 11
16	15,11	26	26/2 = 13

Population mean $= \frac{\text{sum of population}}{\text{Total no. of population}}$

$\mu = \frac{3+7+11+15}{4}$

$\mu = \frac{36}{4}$

$\boxed{\mu = 9}$

Page No.
Date.

2) Sampling Distribution of mean :-

Sample mean ($\bar{x}$)	3	5	7	9	11	13	15	Total
probability	$\frac{1}{16}$	$\frac{2}{16}$	$\frac{3}{16}$	$\frac{4}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	$\frac{1}{16}$	1

mean of sampling Distribution of mean :-

$$E(\bar{x}) = 3 \times \frac{1}{16} + 5 \times \frac{2}{16} + 7 \times \frac{3}{16} + 9 \times \frac{4}{16} + 11 \times \frac{3}{16} + 13 \times \frac{2}{16} + 5 \times \frac{1}{16}$$

$$E(\bar{x}) = \frac{3}{16} + \frac{10}{16} + \frac{21}{16} + \frac{36}{16} + \frac{33}{16} + \frac{26}{16} + \frac{15}{16}$$

$$E(\bar{x}) = \frac{144}{16}$$

$$\boxed{E(\bar{x}) = 9}$$

3) Population standard deviation (S.d)

$\boxed{\mu = 9}$

Variance $\sigma^2 = \frac{\Sigma (x-\mu)^2}{N}$

x	$x-\mu$	$(x-\mu)^2$
3	-6	36
7	-2	4
11	2	4
12	6	36
		80

$$\sigma^2 = \frac{80}{4}$$

$$\boxed{\sigma^2 = 20}$$

S.d = $\sqrt{\text{Variance}}$

S.d = $\sqrt{20}$

$$\boxed{\text{S.d} = 4.4721}$$

4) Standard deviation of sampling distribution of mean.

$$\text{Variance}(\bar{X}) = E(\bar{X})^2 - [E(\bar{X})]^2$$

$$\because E(\bar{X})^2 = 3^2 \times \frac{1}{16} + 5^2 \times \frac{2}{16} + 7^2 \times \frac{3}{16} + 9^2 \times \frac{4}{16} + 11^2 \times \frac{3}{16} + 13^2 \times \frac{2}{16} + 15^2 \times \frac{1}{16}$$

$$E(\bar{X})^2 = \frac{9}{16} + \frac{50}{16} + \frac{147}{16} + \frac{324}{16} + \frac{363}{16} + \frac{338}{16} + \frac{225}{16}$$

$$E(\bar{X})^2 = \frac{1456}{16}$$

$$E(\bar{X})^2 = 91$$

Now,

$$\text{Variance}(\bar{X}) = 91 - (9)^2$$
$$= 91 - 81$$
$$= 10$$

$$S.D(\bar{X}) = \underline{\sqrt{10}}$$

$$S.E(\bar{X}) = 6\bar{X} = \frac{6}{\sqrt{n}} = \frac{\sqrt{20}}{\sqrt{2}}$$
$$= \sqrt{\frac{20}{2}}$$
$$= \underline{\sqrt{10}}$$

Sampling without Replacement $= {}^{N}C_{n}$

Page No.
Date.

Ques- A population consists of four member 3, 7, 11, 15. Consider all possible sample size two which can be drawn without replacement from population. Find the population mean, population standard deviation, the mean of sampling distribution of mean and standard deviation of sampling distribution of mean.

$\Rightarrow$

Sr. No	Sample	total	Sample mean
1	3,7	10	$10/2 = 5$
2	3,11	14	$14/2 = 7$
3	3,5	18	$18/2 = 9$
4	7,11	18	$18/2 = 9$
5	7,15	22	$22/2 = 11$
6	11,15	26	$26/2 = 13$

1) Population mean $= \frac{\text{Sum of Population}}{\text{Total no. of Population}}$

$$(\mu) = \frac{3+7+11+15}{4}$$

$$\boxed{\mu = 9}$$

2) Sampling Distribution of mean :-

Sampling mean ($\bar{x}$)	5	7	9	11	13	Total
probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{2}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Mean of sampling distribution of mean

$$E(\bar{x}) = 5 \times \frac{1}{6} + 7 \times \frac{1}{6} + 9 \times \frac{2}{6} + 11 \times \frac{1}{6} + 13 \times \frac{1}{6}$$

$$E(\bar{x}) = \frac{5}{6} + \frac{7}{6} + \frac{18}{6} + \frac{11}{6} + \frac{13}{6}$$

$$E(\bar{x}) = \frac{54}{6}$$

$$\boxed{E(\bar{x}) = 9}$$

3) Population standard deviation (s.d) is- $\boxed{\mu = 9}$

Variance $\sigma^2 = \frac{\Sigma (x-\mu)^2}{N}$

$$\sigma^2 = \frac{80}{4}$$

$$\sigma^2 = 20$$

x	$(x-\mu)$	$(x-\mu)^2$
3	−6	36
7	−2	4
11	2	4
15	6	36
		= 80

s.d = $\sqrt{\text{variance}}$

$$\boxed{\text{s.d} = \sqrt{20}}$$

4) Standard deviation of sampling distribution of mean

Variance $(\bar{x}) = E(\bar{x})^2 - [E(\bar{x})]^2$

$$\because E(\bar{x})^2 = 5^2 \times \frac{1}{6} + 7^2 \times \frac{1}{6} + 9^2 \times \frac{2}{6} + 11^2 \times \frac{1}{6} + 13^2 \times \frac{1}{6}$$

$$E(\bar{x})^2 = \frac{25}{6} + \frac{49}{6} + \frac{162}{6} + \frac{121}{6} + \frac{169}{6}$$

$$= \frac{526}{6}$$

$E(\bar{x}) = 87.6666$

Variance $= E(\bar{x})^2 - [E(\bar{x})]^2$

$= 87.6666 - 9^2$

$= 87.6666 - 81$

Variance $= 6.6666$

$S.D(\bar{x}) = \sqrt{6.6666} = SE(\bar{x})$

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

where

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

$= \sqrt{\frac{4-2}{4-1}}$

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

$= \sqrt{0.6667}$

$=$

no of sample $= 2$

$\sigma_{\bar{x}} = \frac{\sqrt{20}}{\sqrt{2}} \times \sqrt{\frac{4-2}{4-1}}$

$= \sqrt{\frac{20}{2}} \times \sqrt{\frac{2}{3}}$

$= \sqrt{10} \times \sqrt{\frac{2}{3}}$

$= \sqrt{\frac{20}{3}}$

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

$= \frac{\sqrt{6.6666}}{\sqrt{2}} \times \sqrt{\frac{4-2}{4-1}}$

$= \sqrt{\frac{6.6666}{2}} \times \sqrt{\frac{2}{3}}$

$=$

Page No.
Date.

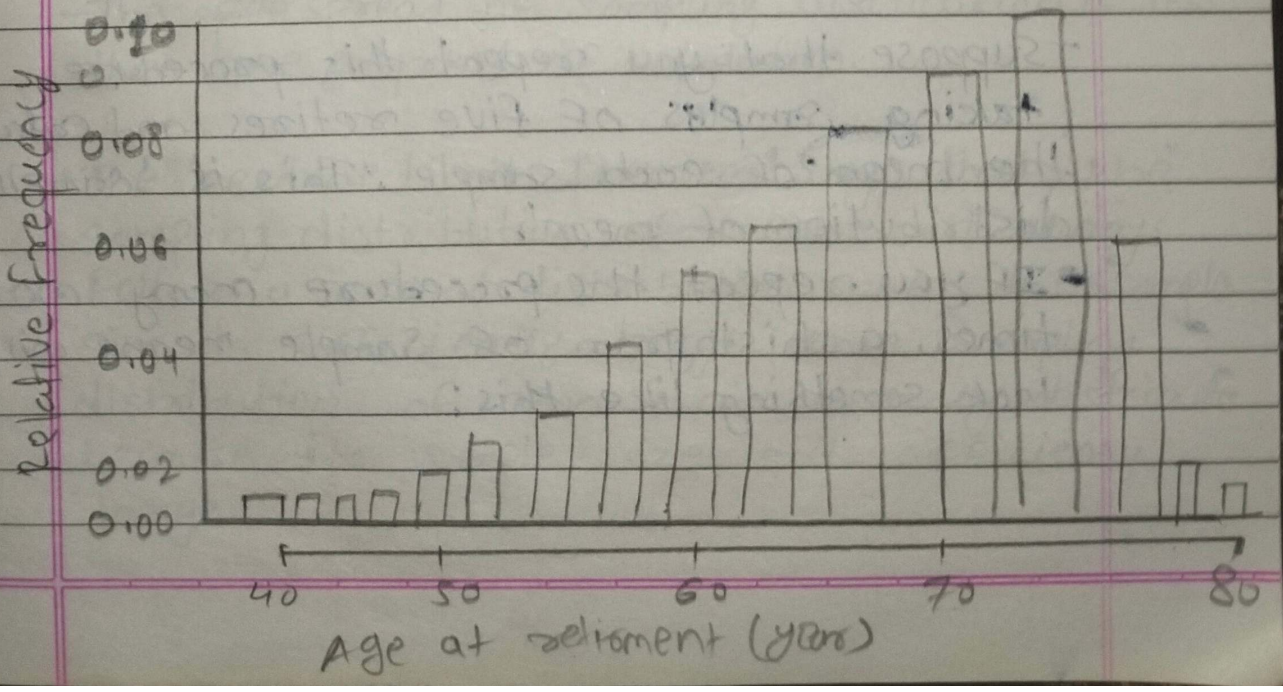
que- With proper example explain sampling distribution of mean and justify it with central limit theorem.

⇒

- central Limith Theorem :-

- The central limit theorem says that the sampling distribution of mean will always be normally distributed as long as sample size is large enough. Regard less of whether the population has normal binomial or any other distribution the sampling distribution of mean will be normal.
- It is one of most fundametal statistical theorems. The "central" in "central limit theorem refers to importance of theorem.

Ex :- Suppose that you're interested in age that people retire in the united state. The population is all retired americans and the distribution of population might look something like this.



- Age at retirement follows a left-skewed distribution. Most people retire within about five years of mean. retirement age of 65 years. However, there's a "long tail" of people who retire much younger, such as at 50 or even 40 years old. The population has standard deviation of 6 years.

- Imagine that you take small sample of the population. you randomly select five retirees and ask then what age they retired.

Ex - central limit theorem:s

sample of n = 5

68, 73, 70, 62, 63

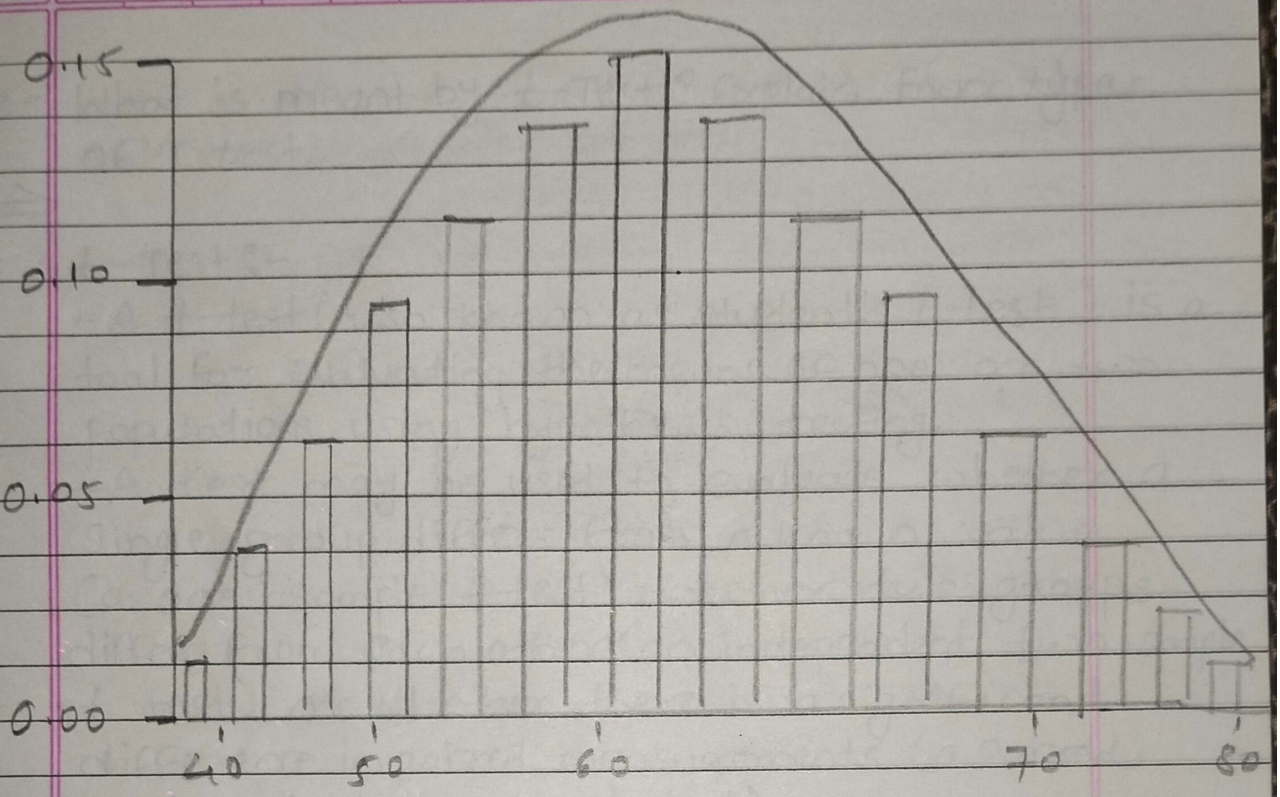
- The mean of sample is an estimate of population mean. It might not be very precise estimate since the sample size is only 5.

∴ mean = (68+73+70+62+63)/5

= 67.2 years.

- Suppose that you repeat this procedure taking samples of five retires and calculating the mean of each sample. This is 'sampling distribution of mean'.

- If you repeat the procedure many more times, a histogram of sample means will look something like this:



Mean age of retirement of small samples.

- Although this sampling distribution is more normally distributed than population it still has a bit of left skew.
- The ~~spee~~ spread of sampling distribution is less than ~~the~~ spread of population.

• The 'central Limit Theorem' says that, the sampling distribution of mean will always follow a normal distribution when the sample size is sufficiently large. This sampling distribution of mean isn't normally distributed because its sample size isn't sufficienty large.