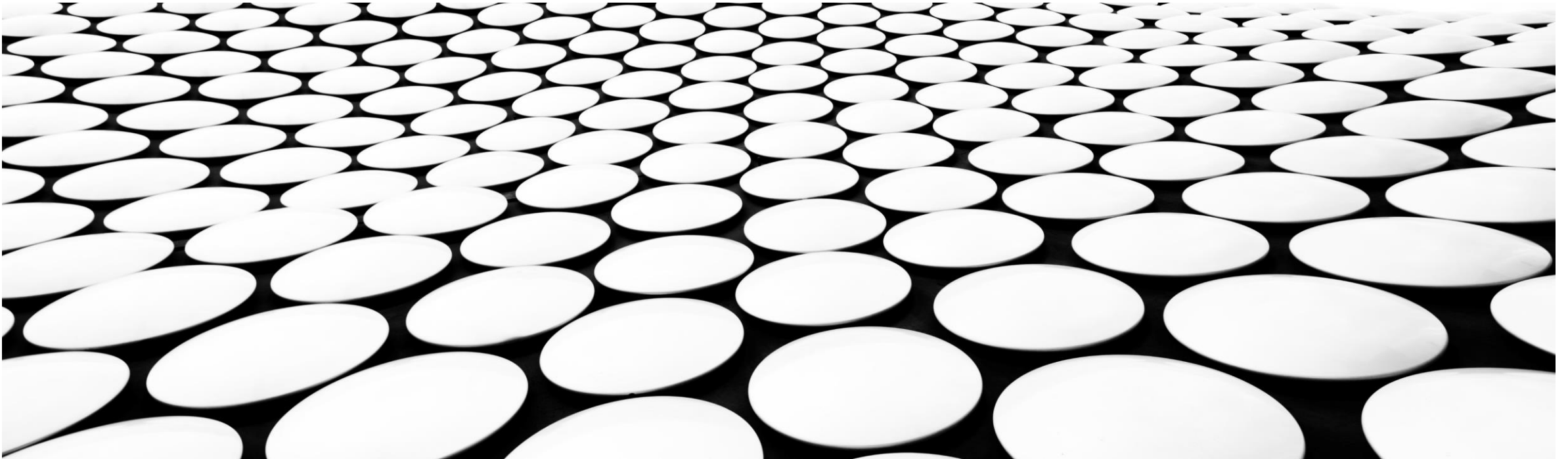
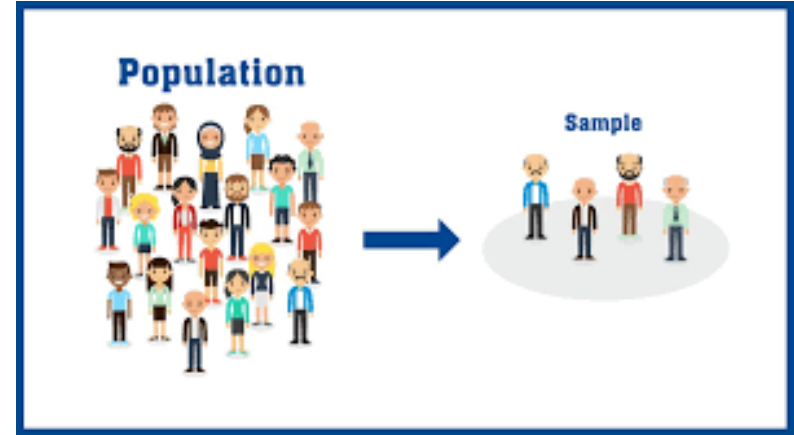


SAMPLING METHODS

BY HIMANSHU SHUKLA (STATISTICAL EXPERT & DATA SCIENTIST)



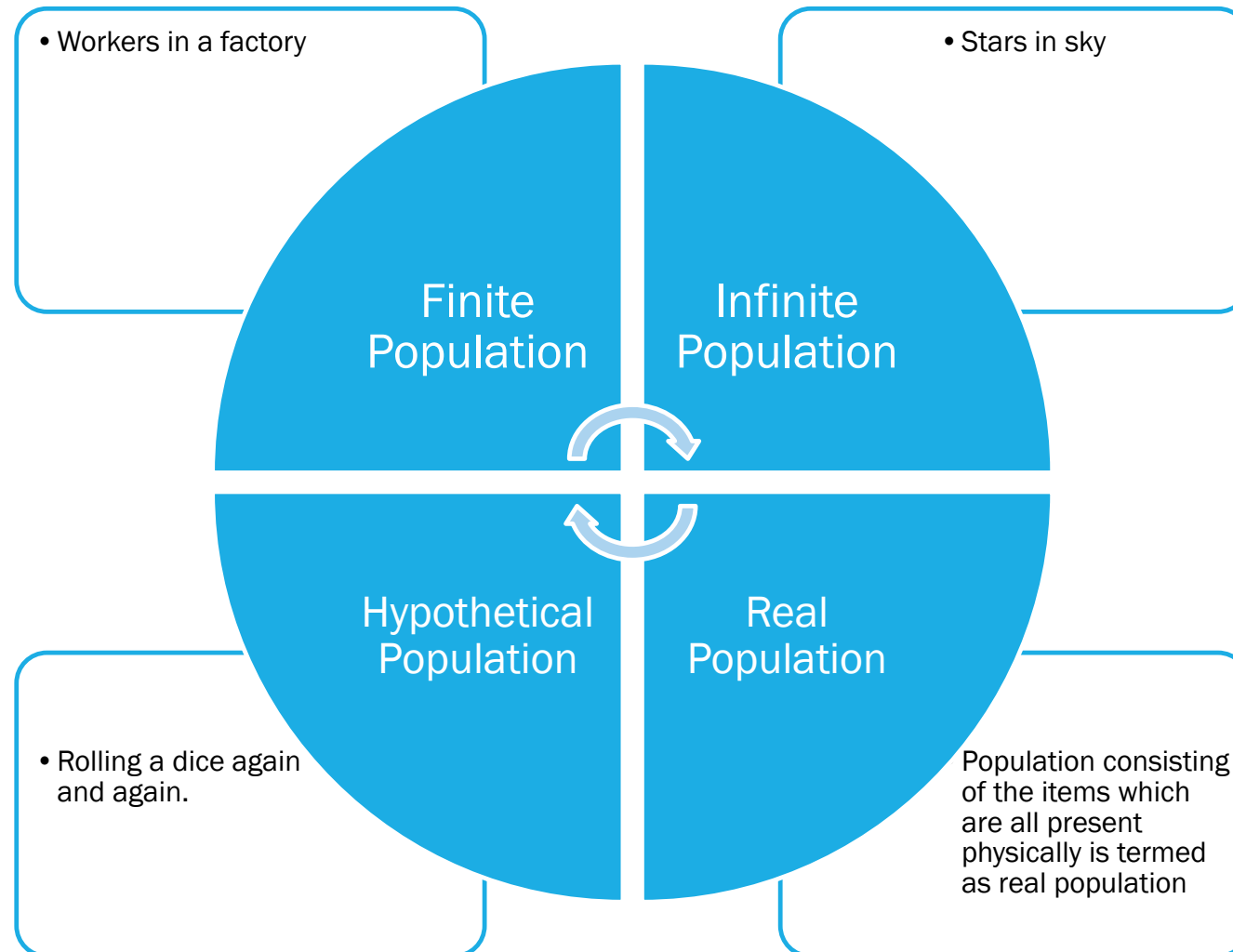


Himanshu Shukla
THE DATA INDUSTRY

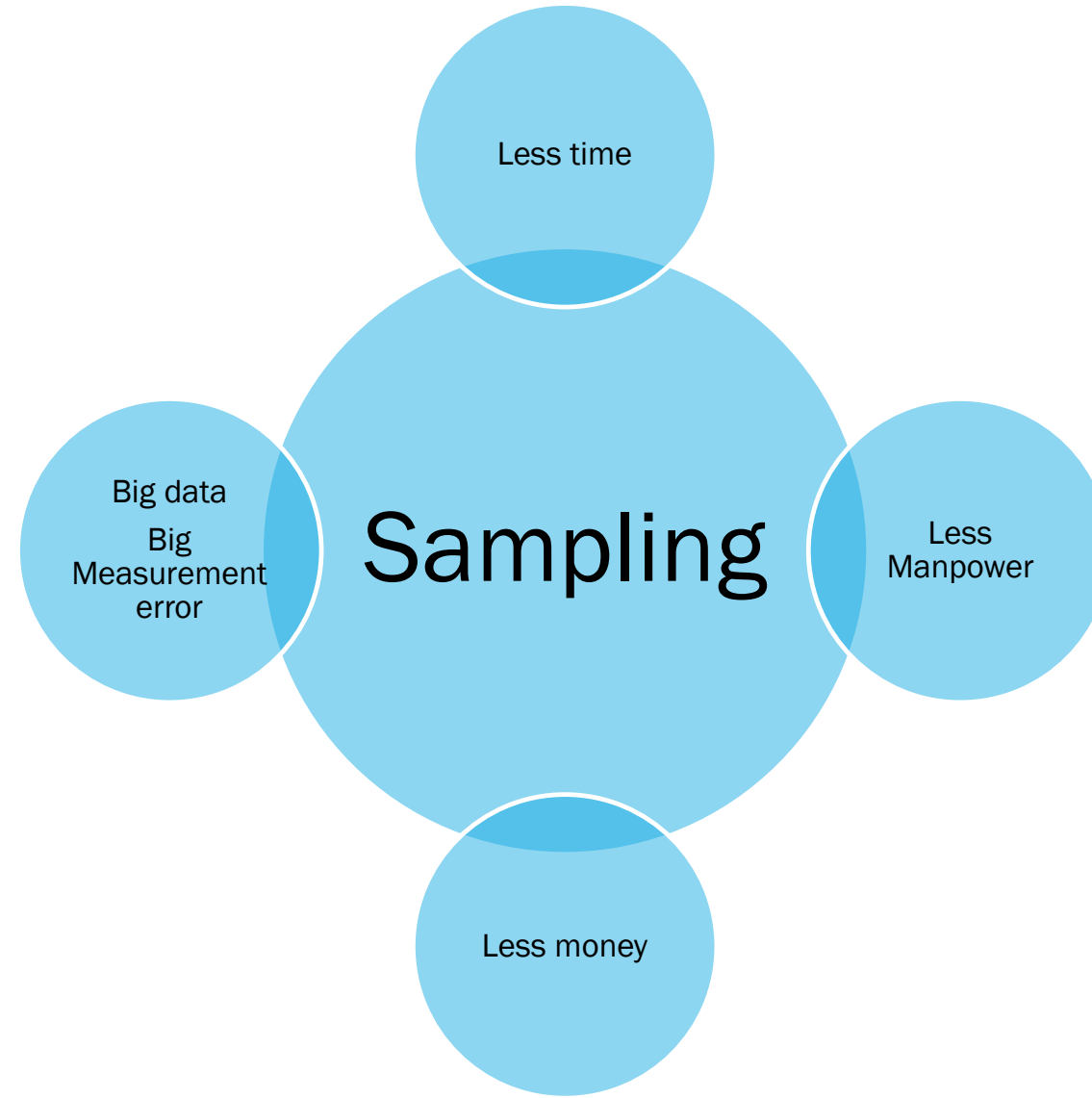
POPULATION IN STATISTICAL SENSE

- One should very well keep in mind that statistical population is **no only the human population** which is usually conceived in literary sense. It is generally a **group or collection of items specified by certain characteristics or defined under certain restrictions**
- A statistical population can be a group of existing objects (e.g. the set of all stars within the Milky Way galaxy) or a hypothetical and potentially infinite group of objects conceived as a generalization from experience (e.g. the set of all possible hands in a game of poker).

TYPES OF POPULATION



WHAT IS THE NEED OF SAMPLING AS COMPARED TO COMPLETE ENUMERATION

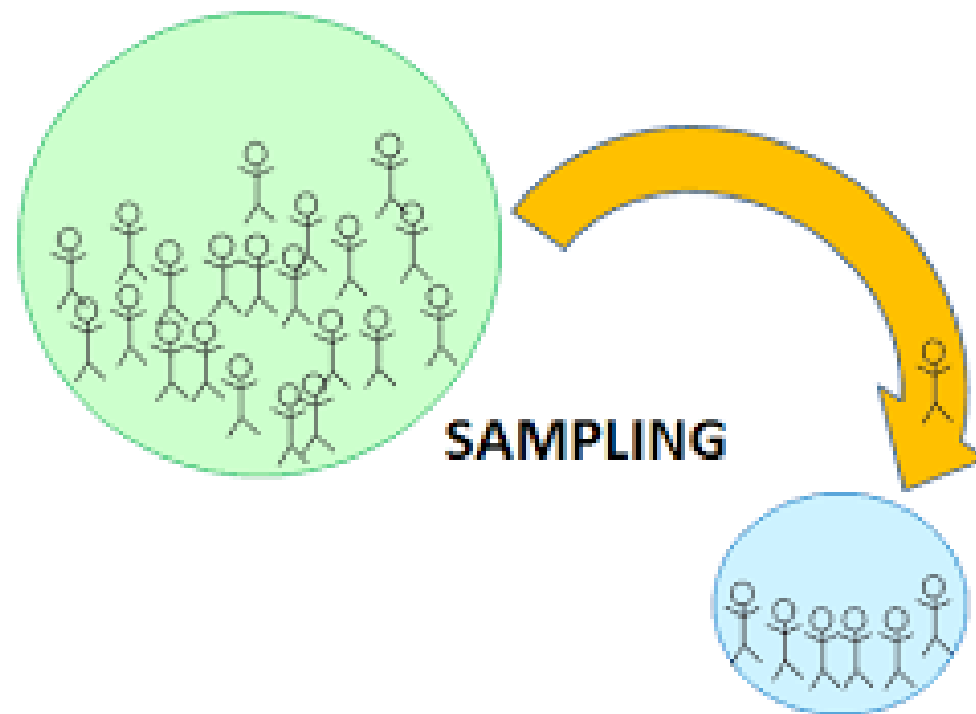


WHAT IS SAMPLE

- Sample is a part or fraction of population selected on some basis. But keep in mind its principle is it should be true representative of the population.

Two Main Questions

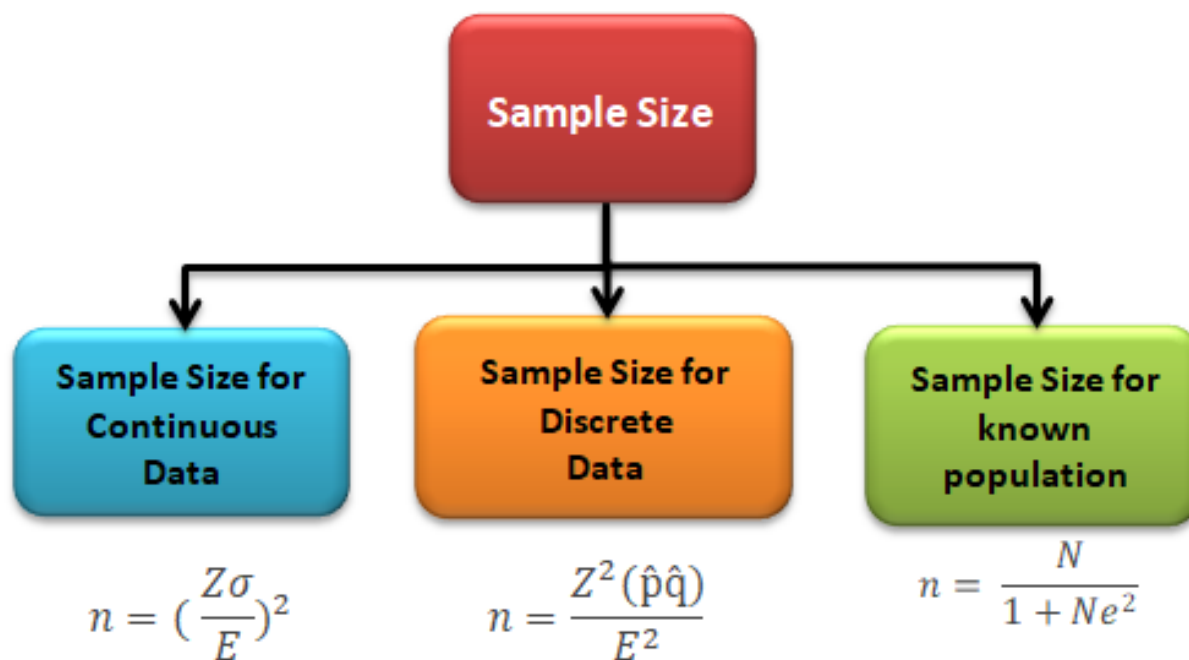
- Which technique or method we adopt?
- What should be size of sample?



SAMPLE SIZE

- Sample size =
$$\frac{\frac{z^2 \times p(1-p)}{e^2}}{1 + \left(\frac{z^2 \times p(1-p)}{e^2 N}\right)}$$

- N = Population Size
- z = Z score
- e = margin of error
- p = standard deviation

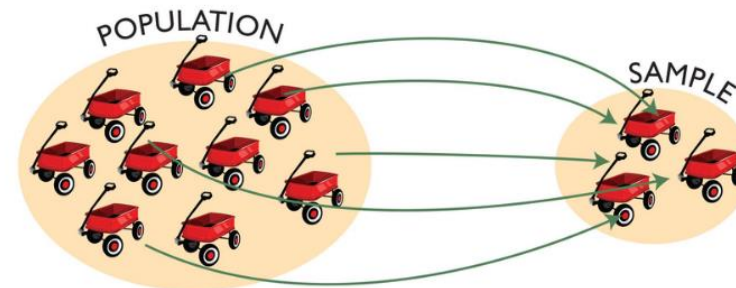


POPULATION

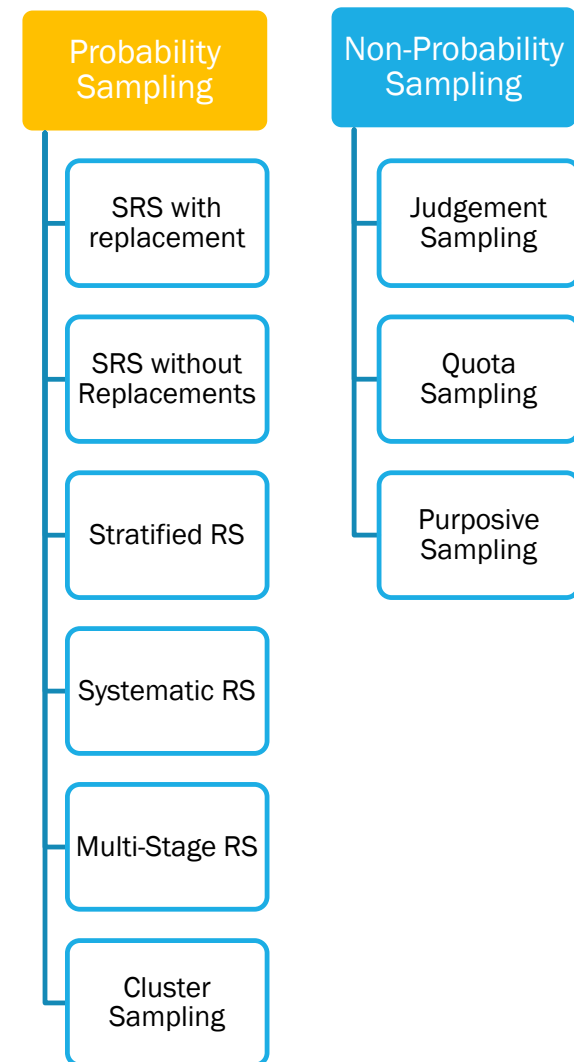
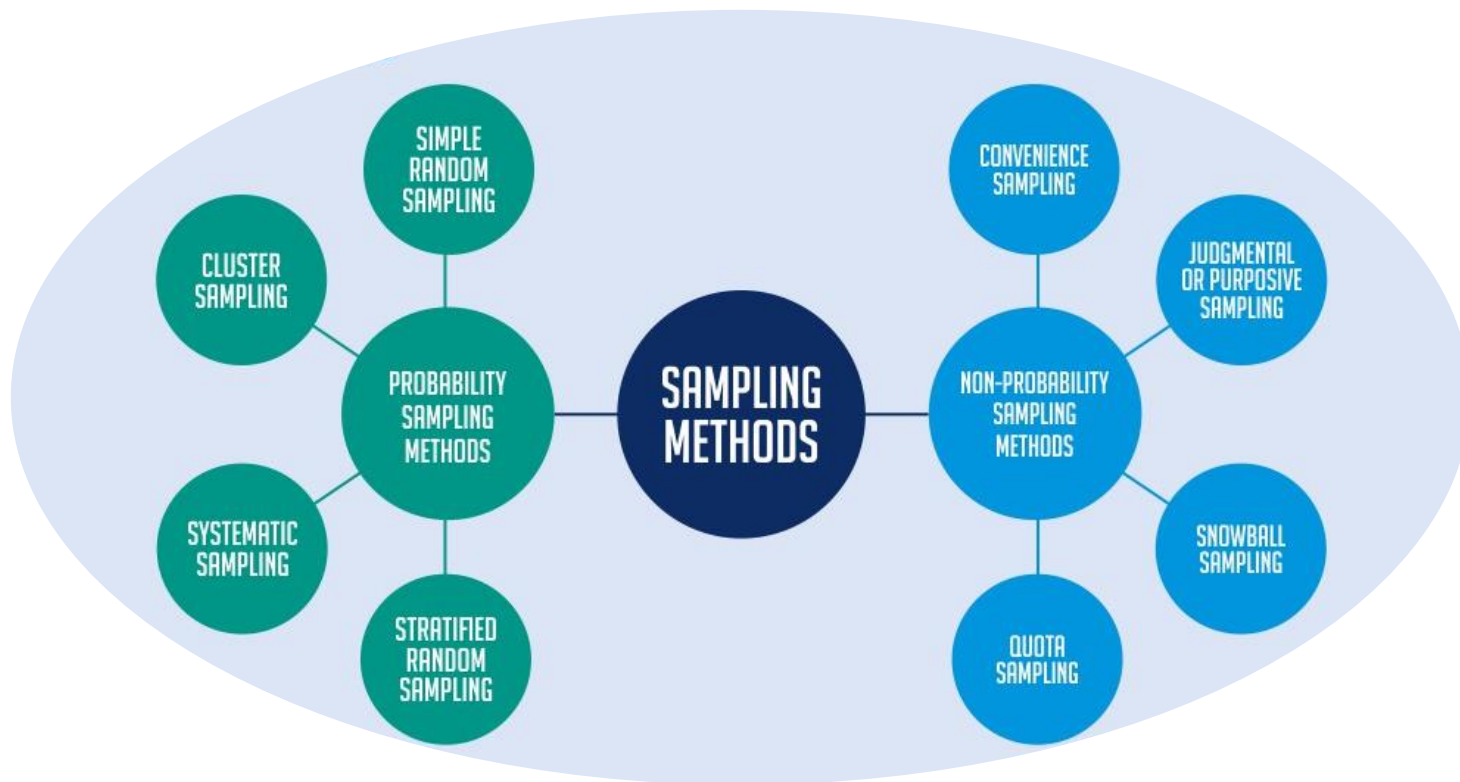
- The measurable quality is called a parameter.
- The population is a complete set.
- Reports are a true representation of opinion.
- It contains all members of a specified group.

SAMPLE

- The measurable quality is called a statistic.
- The sample is a subset of the population.
- Reports have a margin of error and confidence interval.
- It is a subset that represents the entire population.



	Sample Size	Mean	Standard Deviation	Variance
Population	N	μ	σ	σ^2
Sample	n	\bar{x}	s	s^2



NON-PROBABILITY SAMPLING

Judgmental sampling / Purposive Sampling

- Judgmental sampling is a non-probability sampling technique where the researcher selects units to be sampled based on their knowledge and professional judgment.
- It is also known as purposive sampling and authoritative sampling.

Quota Sampling

- Quota sampling is where you take a very tailored sample that's in proportion to some characteristic or trait of a population.

SIMPLE RANDOM SAMPLING

- When equal probability of selection is attached to each sampling unit at each draw, the selection procedure is known as random sampling.

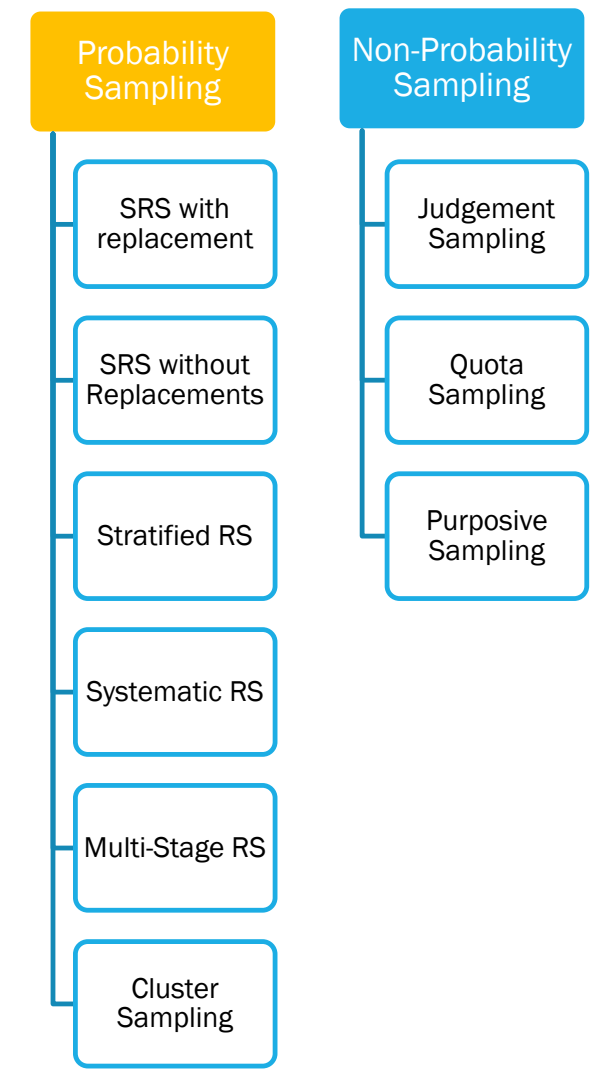
Two approaches aim to minimize any biases in the process of simple random sampling:

•Method of lottery

Using the lottery method is one of the oldest ways and is a mechanical example of random sampling. In this method, the researcher gives each member of the population a number. Researchers draw numbers from the box randomly to choose samples.

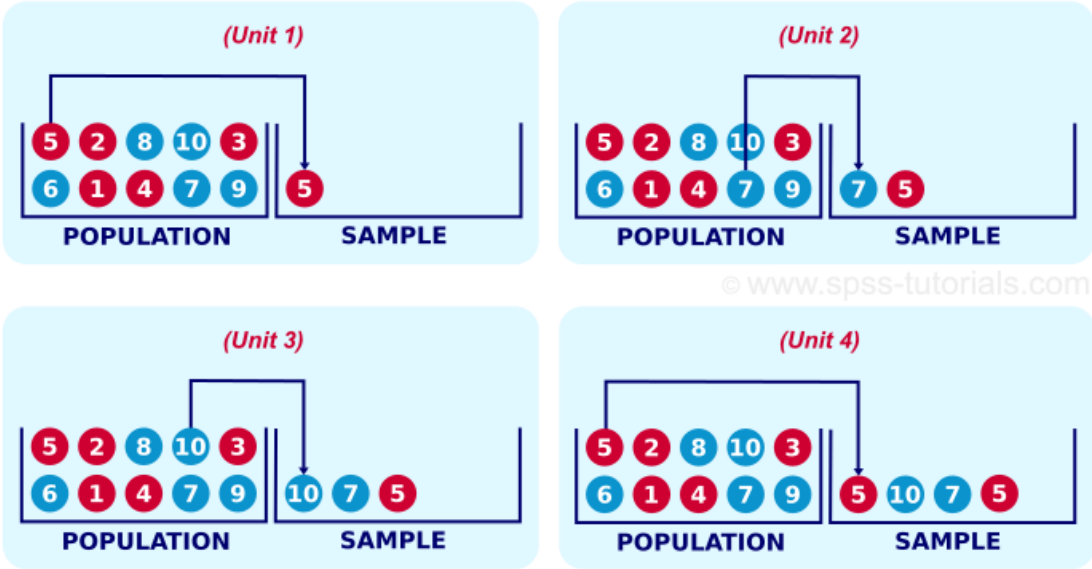
•Use of random numbers

The use of random numbers is an alternative method that also involves numbering the population. The use of a number table similar to the one below can help with this sampling technique.

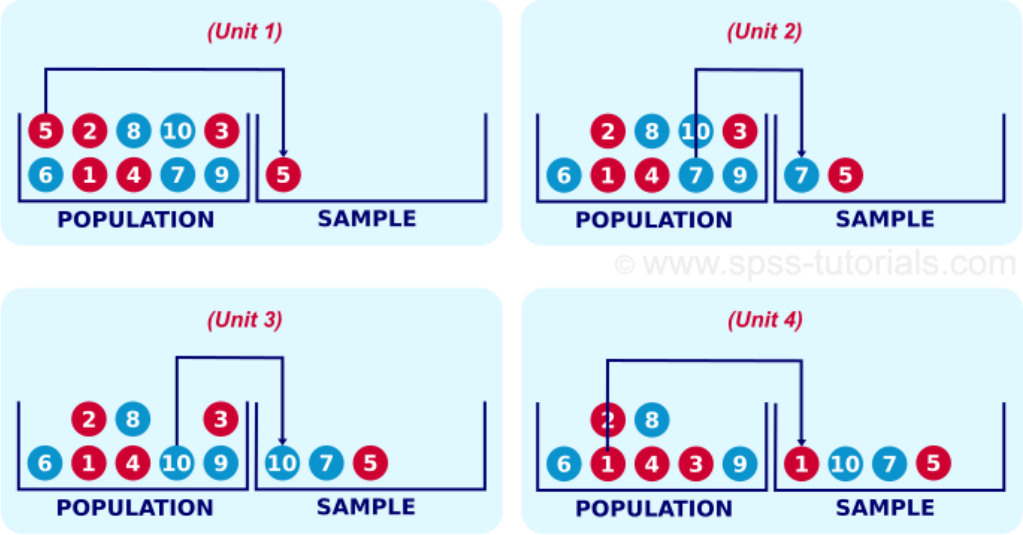


WITH REPLACEMENT MEANS THE SAME ITEM CAN BE CHOSEN MORE THAN ONCE. WITHOUT REPLACEMENT MEANS THE SAME ITEM CANNOT BE SELECTED MORE THAN ONCE.

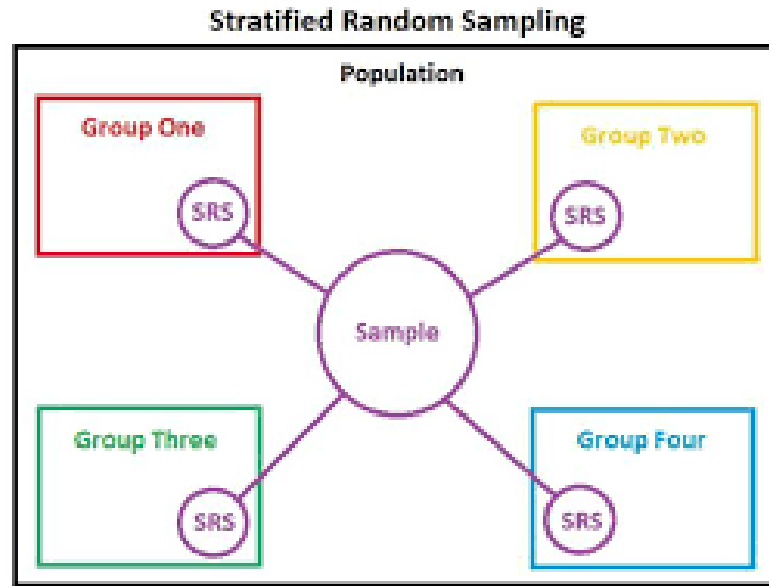
SIMPLE RANDOM SAMPLING *WITH* REPLACEMENT



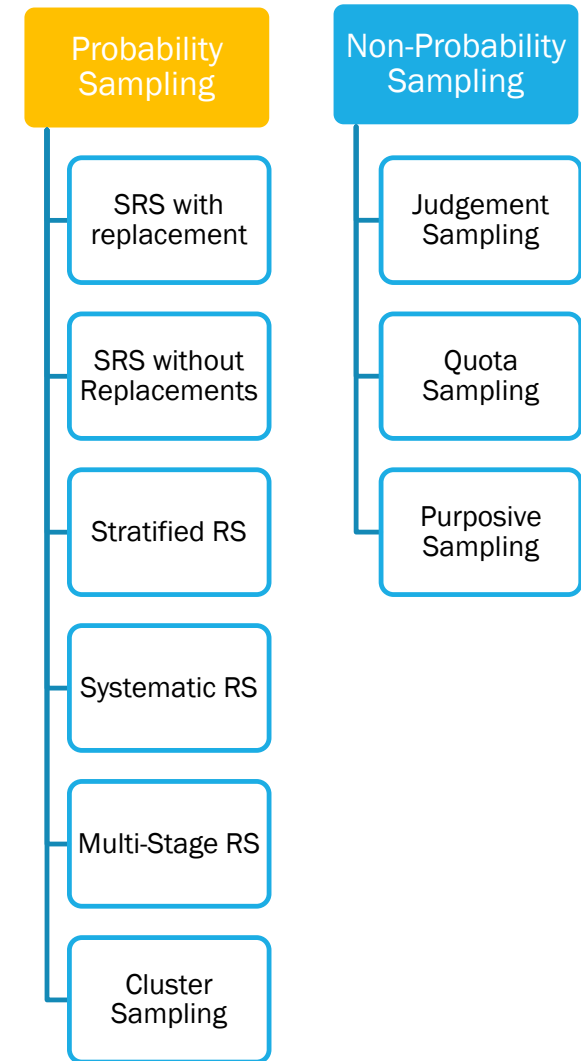
SIMPLE RANDOM SAMPLING *WITHOUT* REPLACEMENT



STRATIFIED RANDOM SAMPLING

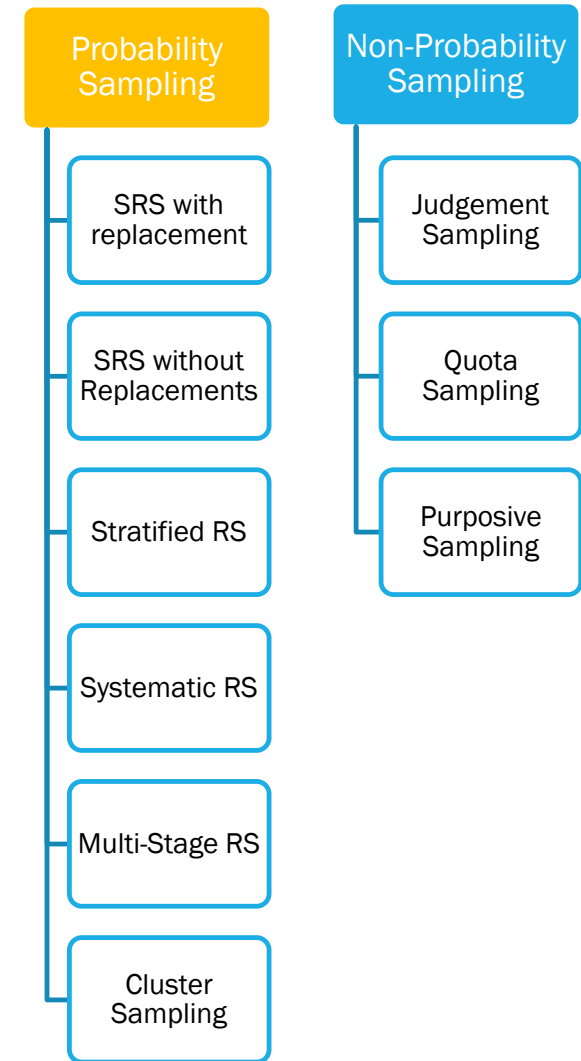


- In this method we did sampling in two steps:
 - First the whole population is divided into **homogenous** group under certain criterion.(which is called strata)
 - Second, sample is drawn **randomly from each stratum** independently.
- This sampling comes under the category of restricted sampling.
- Strata are formed on some basis such as localities in a city, districts in a state, etc.



SYSTEMATIC SAMPLING

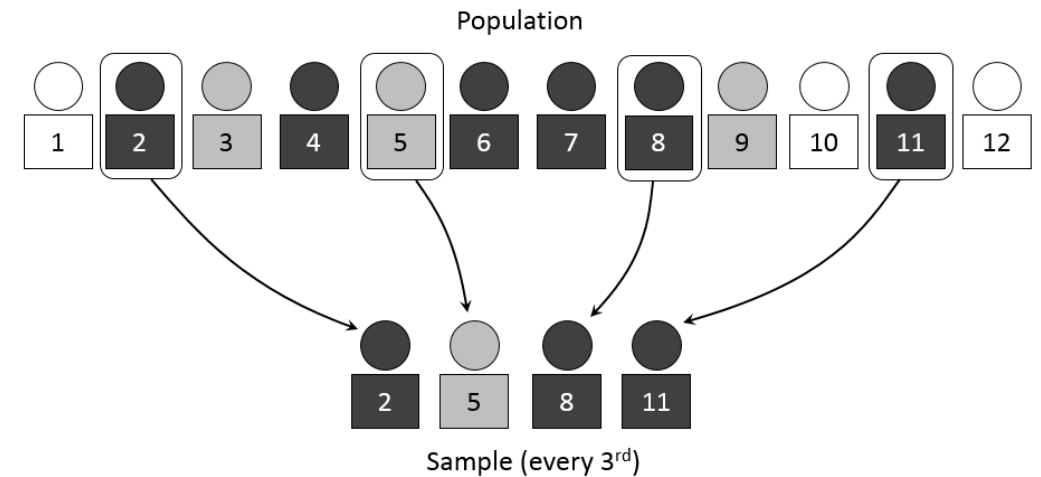
- Systematic sampling is a **probability sampling method where researchers select members of the population at a regular interval** – for example, by selecting every 15th person on a list of the population. If the population is in a random order, this can imitate the benefits of simple random sampling.



Systematic Sampling

In order to take a **systematic sample**:

- 1 Order the population and give each data entry a unique reference number.
- 2 Calculate the number of items of data in the sample.
- 3 Calculate the interval.
- 4 Use a random number generator to select the first item of data.
- 5 Select the remaining items of data following the given sequence.

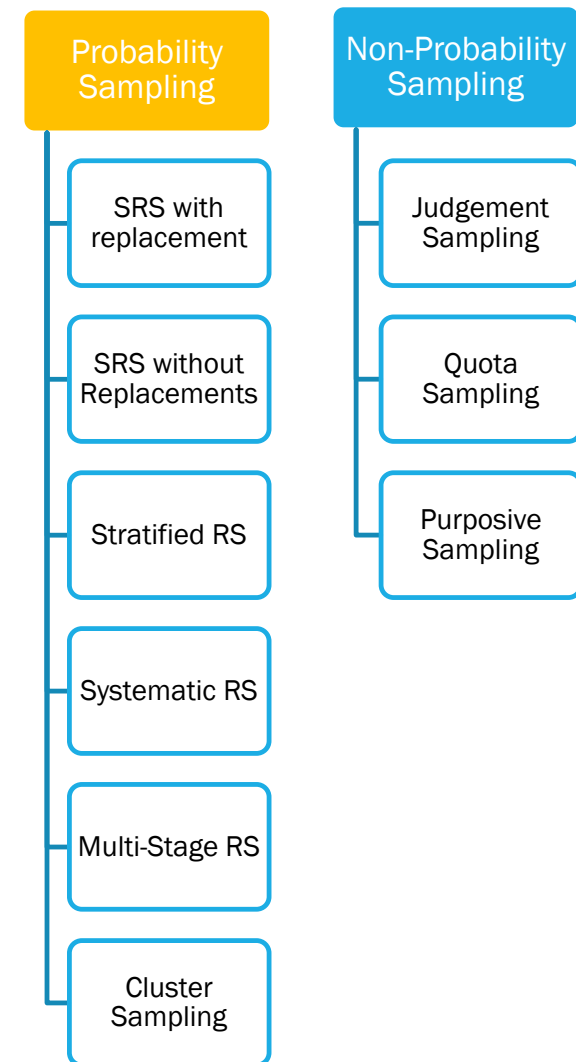
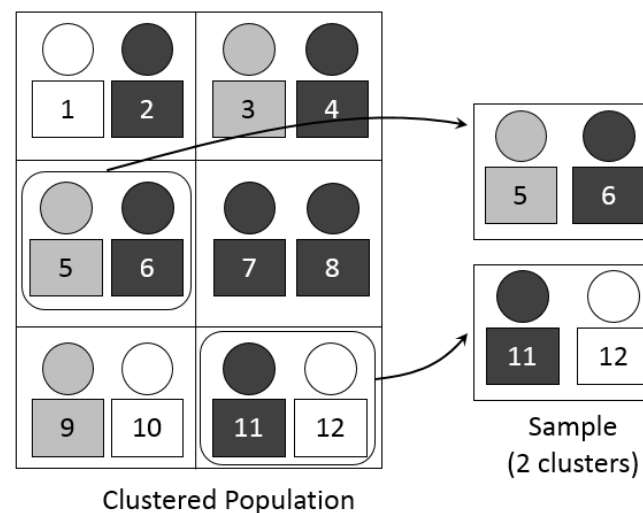


CLUSTER SAMPLING

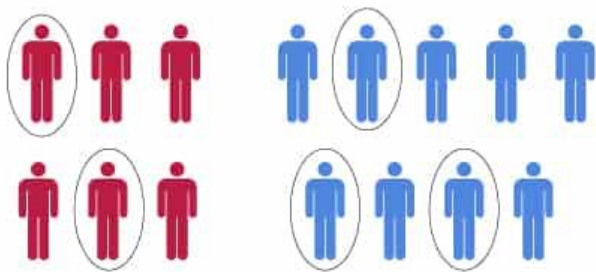
- Cluster sampling is a probability sampling method in which you divide a population into clusters, such as districts or schools, and then randomly select some of these clusters as your sample.

You thus decide to use the cluster sampling method.

- Step 1: Define your population. ...
- Step 2: Divide your sample into clusters. ...
- Step 3: Randomly select clusters to use as your sample. ... Usually SRSWOR is used to select clusters or area .But can other design be also use
- Step 4: Collect data from the sample.

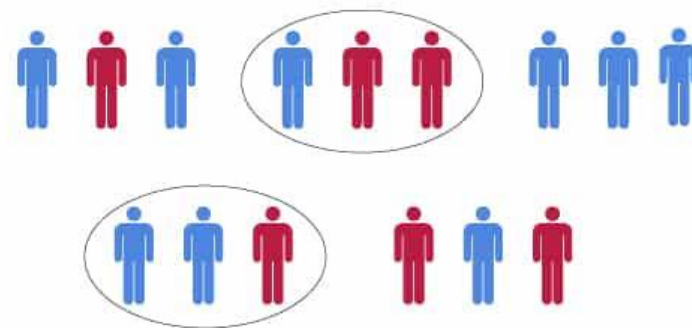


Stratified random sampling



VS

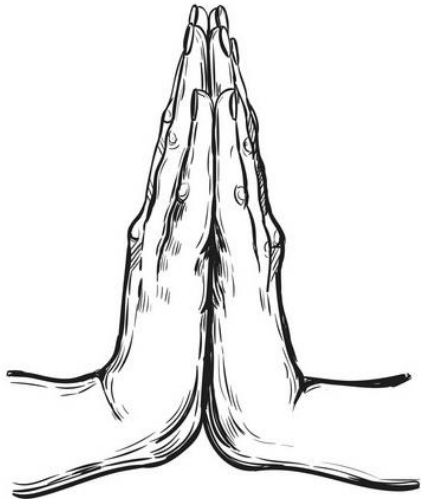
Cluster sampling



MERITS AND DEMERITS OF SAMPLING DESIGNS

Design	Merits	Demerits
<u>PROBABILITY SAMPLING</u> Simple Random Sampling	<ol style="list-style-type: none"> Selection is entirely on chance No probability of personal bias Represents the population in a better way Relatively easy to assess the accuracy of this estimate 	<ol style="list-style-type: none"> Certain strata may be over or under represented in the sample Needs complete cataloguing of the universe Size of sample is relatively larger for enhancing reliability Cases often are distributed too widely making it time taking and costly
Stratified Sampling	<ol style="list-style-type: none"> More representative Greater accuracy Greater geographical concentration 	<ol style="list-style-type: none"> Needs utmost care while dividing the population into distinct strata Each item within the stratum needs random selection
Systematic Sampling	<ol style="list-style-type: none"> Simple and convenient Time and work involved in sampling are relatively small 	<ol style="list-style-type: none"> Becomes less representative if we are dealing with hidden periodicity
Cluster Sampling	<ol style="list-style-type: none"> It allows large coverage and fieldwork in a relatively concentrated area Useful when we even do not know the complete population frame 	<ol style="list-style-type: none"> All cases must be included from the sampled cluster Sampling error in a cluster sample is much greater than in true random sample

The END



Namaste

THANKS!

Ab Kuch Puchna Mat, Thoda Khud Bhi padho or Google karo.

You can find me at:

dataindustryhimanshu@gmail.com