



# SAMPLING TECHNIQUES

DR.BHARAT PAUL

# CONTENTS

- Introduction
- Need for sampling
- Sampling Process
- Essentials of Sampling
- Methods of Sampling
  - Non Probability Sampling
  - Probability Sampling
- Errors in Sampling
- References



# INTRODUCTION

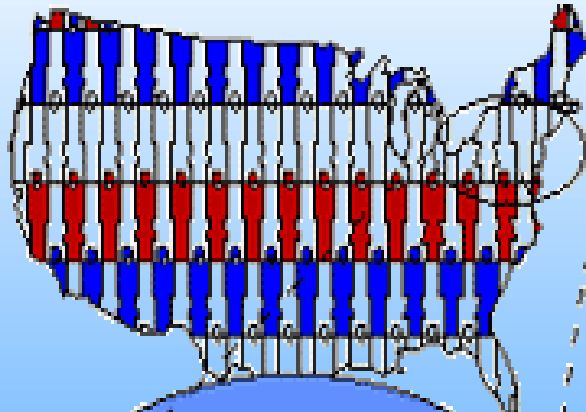
- **Population/Universe:** in statistics denotes the aggregate from which sample (items) is to be taken.
- A population can be defined as including all people or items with the characteristic one wishes to understand.
- Because there is very rarely enough time or money to gather information from everyone or everything in a population, the goal becomes finding a representative sample (or subset) of that population.

# INTRODUCTION

- **Sampling frame** is the list from which the potential respondents are drawn .
- **A sample** is “a smaller (but hopefully representative) collection of units from a population used to determine truths about that population” (Field, 2005)

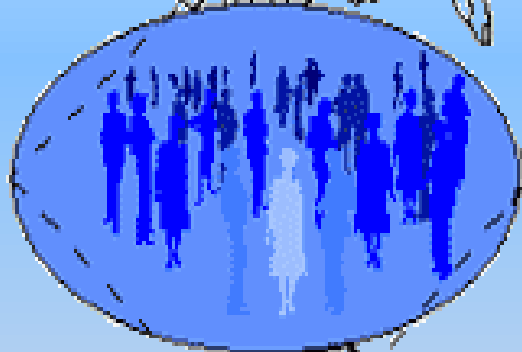
# SAMPLING BREAKDOWN

Who do you want to generalize to?



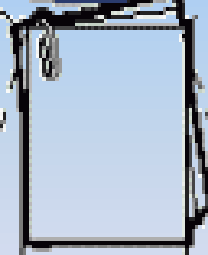
The Theoretical Population

What population can you get access to?



The Study Population

How can you get access to them?



The Sampling Frame

Who is in your study?

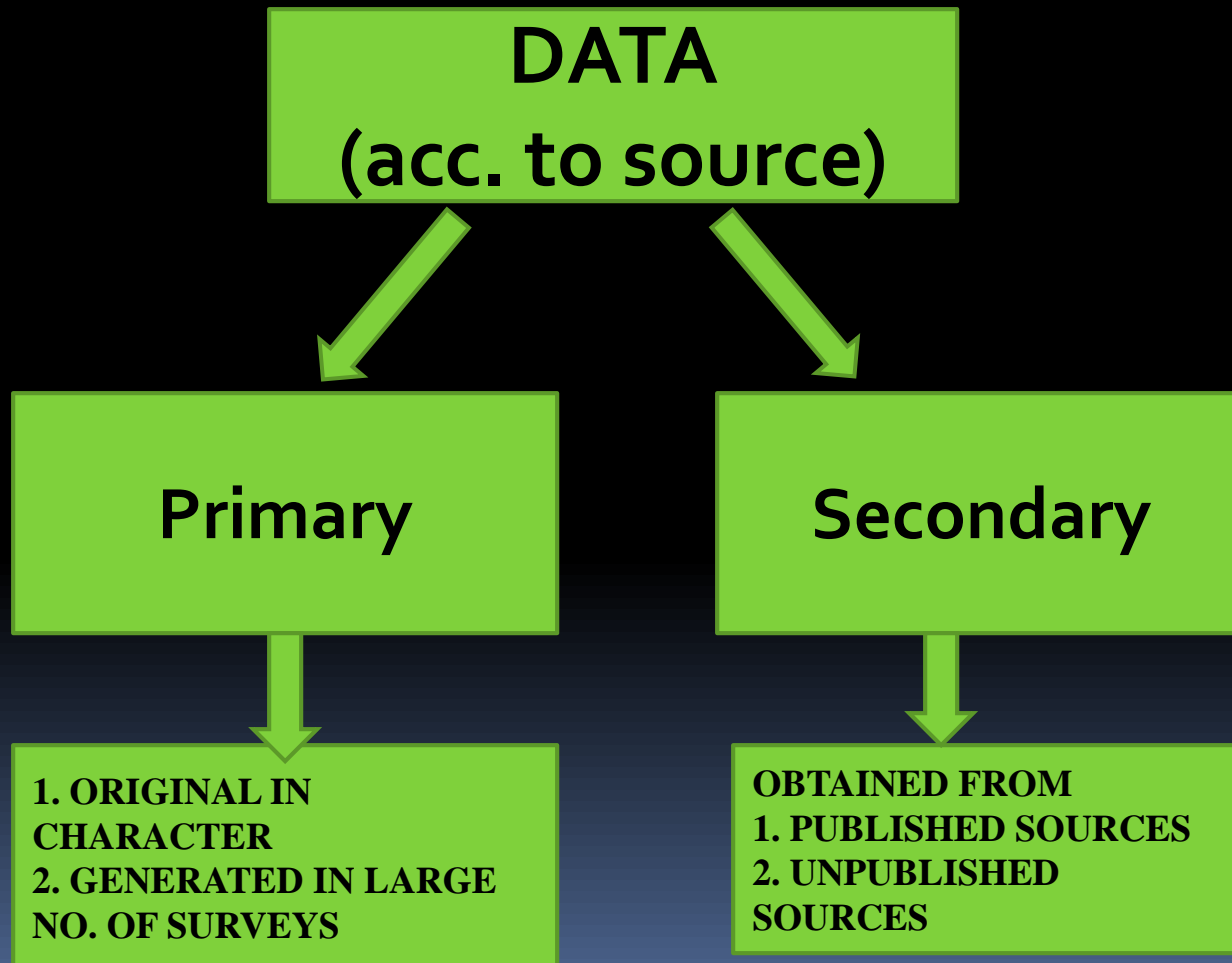


The Sample

# SAMPLING

- **Sampling:** the process of learning about population on the basis of sample drawn from it.
- **Three elements** in process of sampling:
  - Selecting the sample
  - Collecting the information
  - Making inference about population
- **Statistics:** values obtained from study of a sample.
- **Parameters:** such values from study of population.

# NEED FOR SAMPLING



# NEED FOR SAMPLING

- When secondary data are not available for the problem under study , primary data is collected.
- Two methods –
  - **Census method** or complete enumeration method
  - **Sample method**



# CENSUS (Complete Enumeration Survey)

- **Merits**


- Data obtained from each and every unit of population.
- Results: more representative, accurate, reliable.
- Basis of various surveys.

- **Demerits**

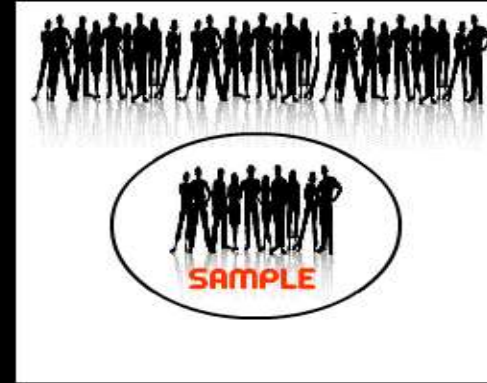
- More effort ,money , time.
- Big problem in underdeveloped countries.



# ADVANTAGES OF SAMPLING

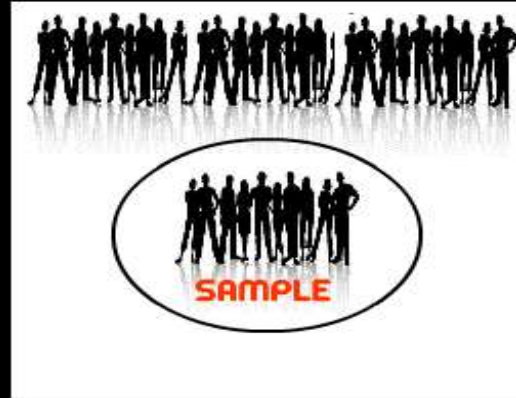
- Less resources (time, money)
  - Less workload.
  - Gives results with known accuracy that can be calculated mathematically.
- 

# THEORETICAL BASIS OF SAMPLING



- On the basis of sample study we can predict and generalize the **behavior of mass phenomena.**
- There is no statistical population whose elements would vary from each other without limit.

# THEORETICAL BASIS OF SAMPLING



- **Law of Statistical Regularity-**
  - Sample is taken at random from a population, it is likely to possess same characteristics as that of population.
- **Law of inertia of large numbers-**
  - Larger the size of sample, more accurate the results are likely to be.

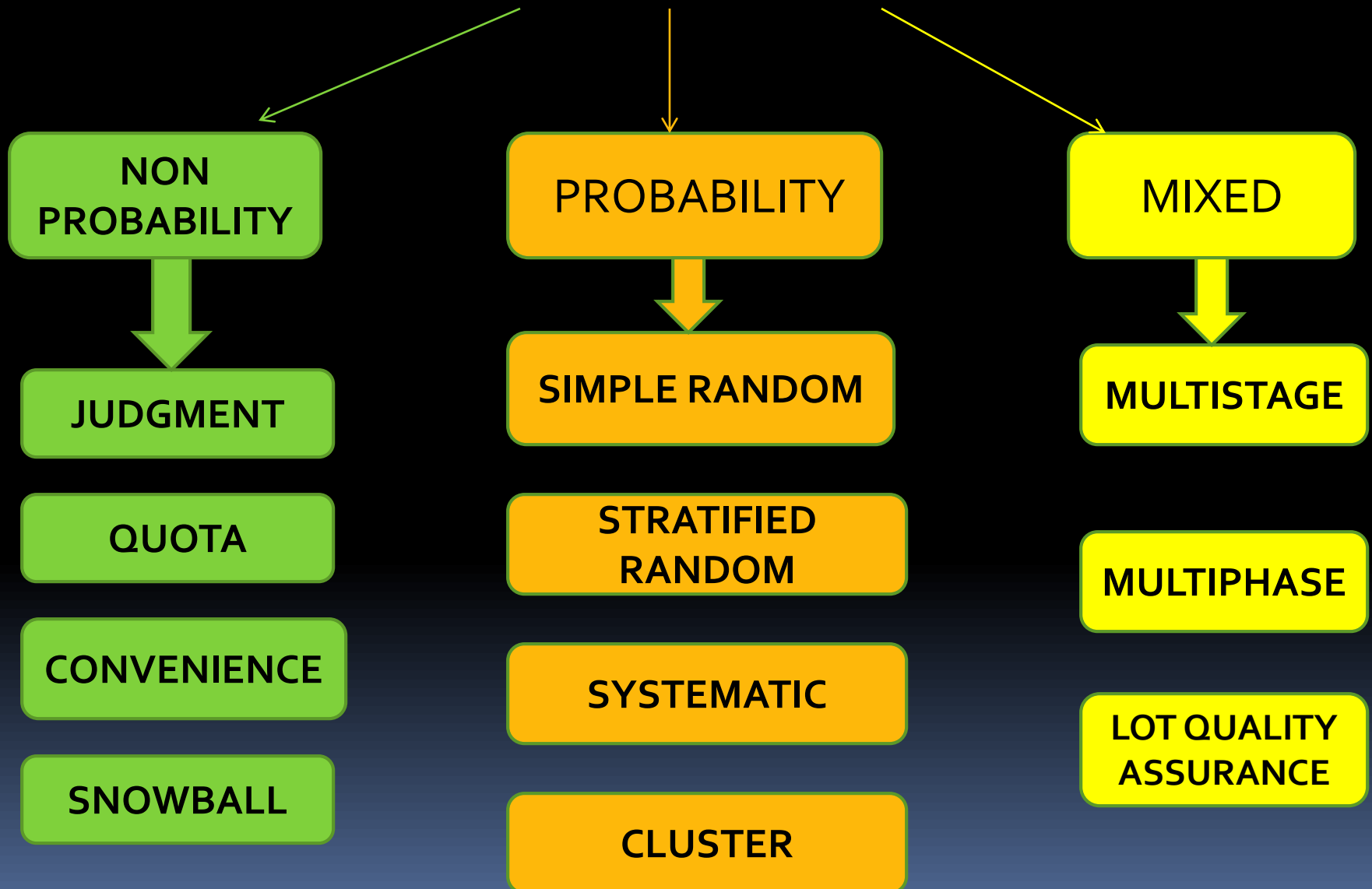
# SAMPLING PROCESS

- Defining the **population of concern**.
- Specifying a **sampling frame**, a set of items or events possible to measure.
- Specifying a **sampling method** for selecting items or events from the frame.
- Determining the **sample size**.
- Implementing the sampling plan.
- Sampling and data collection

# ESSENTIALS OF SAMPLING

- **Representativeness** - ensure by random selection
- **Adequacy** - sample size
- **Independence** - same chance of selection
- **Homogeneity** - no basic difference in nature of units.

# SAMPLING METHODS





# NON PROBABILITY SAMPLING



# JUDGMENT SAMPLING

- Judgment/Purposive/Deliberate sampling.
- Depends exclusively on the **judgment of investigator.**
- Sample selected which investigator thinks to be most **typical of the universe.**

# JUDGMENT SAMPLING

- **Merits**

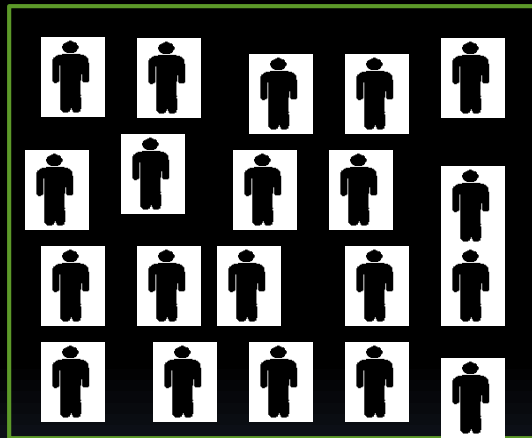
- Small no. of sampling units
- Study unknown traits/case sampling
- Urgent public policy & business decisions

- **Demerits**

- Personal prejudice & bias
- No objective way of evaluating reliability of results

# JUDGMENT SAMPLING - EXAMPLE

Sample size for a study=8



CLASS OF 20 STUDENTS

JUDGMENT



SAMPLE OF 8  
STUDENTS

# CONVENIENCE SAMPLING

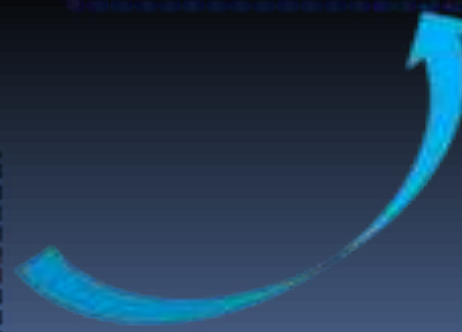
- **Convenient** sample units selected.
- Selected neither by probability nor by judgment.
- **Merit** – useful in pilot studies.
- **Demerit** – results usually biased and unsatisfactory.

# CONVENIENCE SAMPLING - EXAMPLE

Class of 100 students



20 Students selected as per convenience

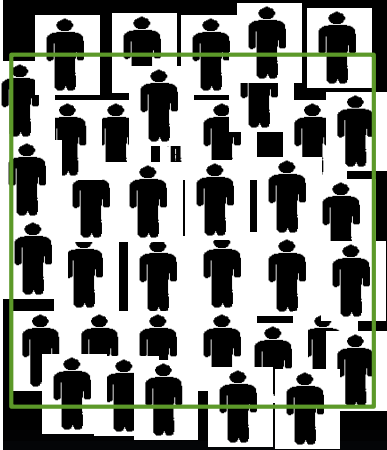


# QUOTA SAMPLING

- Most commonly used in non probability sampling.
- Quotas set up according to some specified characteristic.
- Within the quota , selection depends on personal judgment.
- **Merit**- Used in public opinion studies
- **Demerit** – personal prejudice and bias

# QUOTA SAMPLING - EXAMPLE

Radio listening survey



Quota Formation

60%  
housewives

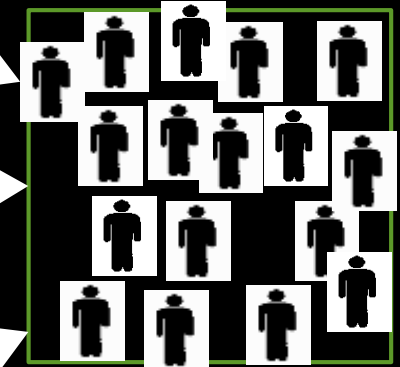
25%  
farmers

15% children  
under age 15

300

125  
Personal  
judgement

75



Interview 500 people

500 people

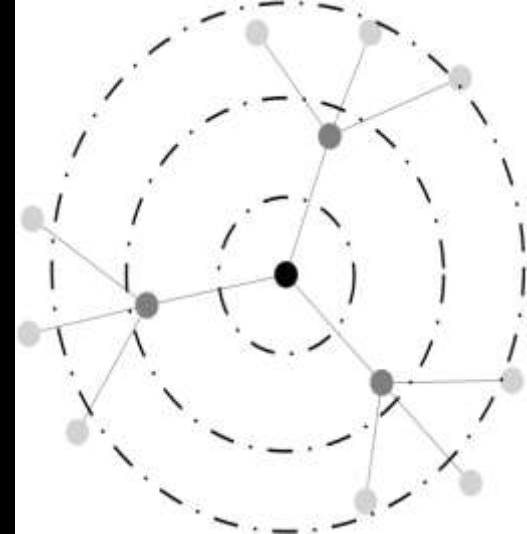
# SNOWBALL SAMPLING

- A special non probability method used when the desired **sample characteristic is rare**.
- It may be extremely difficult or cost prohibitive to locate respondents in these situations.
- Snowball sampling **relies on referrals** from initial subjects to generate additional subjects.



# SNOWBALL SAMPLING - STEPS

- Make **contact with one or two cases** in the population.
- Ask **these cases to identify further cases.**
- Ask these new cases to identify further new cases.
- Stop when either no new cases are given or the sample is as large as is manageable.



# SNOWBALL SAMPLING

- **Merit**

- access to difficult to reach populations (other methods may not yield any results).

- **Demerit**


- not representative of the population and will result in a biased sample as it is self-selecting.



# PROBABILITY SAMPLING



# SIMPLE RANDOM SAMPLING

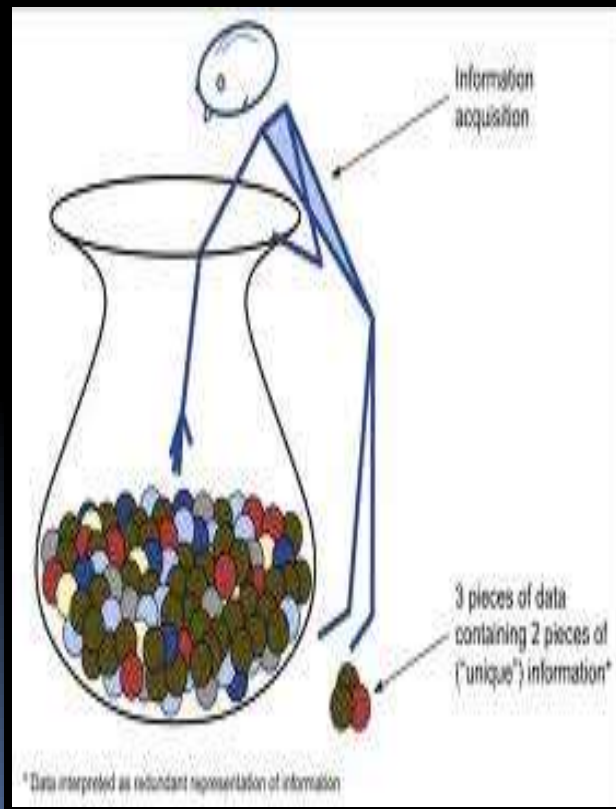
- Each unit has an **equal opportunity** of being selected.
  - **Chance** determines which items shall be included.
- 

# SIMPLE RANDOM SAMPLING

- The sample is a simple random sample if any of the following is true (Chou) –
  - All items selected independently.
  - At each selection , all remaining items have same chance of being selected.
  - All the possible samples of a given size are equally likely to be selected.

# Simple or unrestricted random sampling

## Lottery method



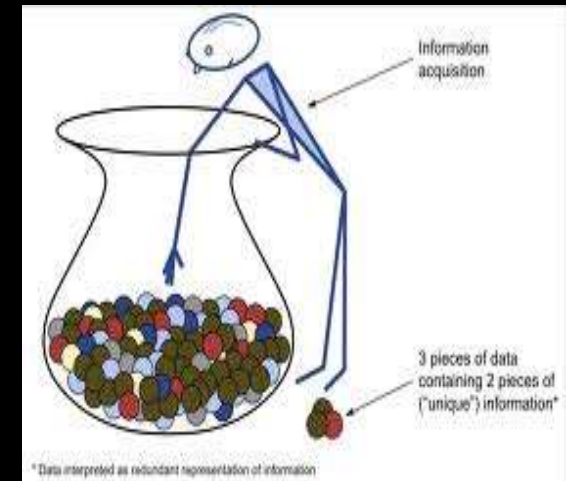
## Random number tables

### Part of a Table of Random Numbers

|       |       |       |       |
|-------|-------|-------|-------|
| 61424 | 20419 | 86546 | 00517 |
| 90222 | 27993 | 04952 | 66762 |
| 50349 | 71146 | 97668 | 86523 |
| 85676 | 10005 | 08216 | 25906 |
| 02429 | 19761 | 15370 | 43882 |
| 90519 | 61988 | 40164 | 15815 |
| 20631 | 88967 | 19660 | 89624 |
| 89990 | 78733 | 16447 | 27932 |

# LOTTERY METHOD -

- **With replacement-**
  - Probability each item:  $1/N$
- **Without replacement –**
  - **Probability 1<sup>st</sup> draw:  $1/N$**
  - **Probability 2<sup>nd</sup> draw:  $1/N-1$**



# TABLE OF RANDOM NUMBERS

|       |       |       |       |
|-------|-------|-------|-------|
| 73735 | 45963 | 78134 | 63873 |
| 02965 | 58303 | 90708 | 20025 |
| 98859 | 23851 | 27965 | 62394 |
| 33666 | 62570 | 64775 | 78428 |
| 81666 | 26440 | 20422 | 05720 |

|       |       |       |       |
|-------|-------|-------|-------|
| 15838 | 47174 | 76866 | 14330 |
| 89793 | 34378 | 08730 | 56522 |
| 78155 | 22466 | 81978 | 57323 |
| 16381 | 66207 | 11698 | 99314 |
| 75002 | 80827 | 53867 | 37797 |

|       |       |       |       |
|-------|-------|-------|-------|
| 99982 | 27601 | 62686 | 44711 |
| 84543 | 87442 | 50033 | 14021 |
| 77757 | 54043 | 46176 | 42391 |
| 80871 | 32792 | 87989 | 72248 |
| 30500 | 28220 | 12444 | 71840 |



# SIMPLE RANDOM SAMPLING

- **Merits**


- No personal bias.
- Sample more representative of population.
- Accuracy can be assessed as sampling errors follow principals of chance.

- **Demerits**

- Requires completely catalogued universe.
- Cases too widely dispersed - more time and cost.

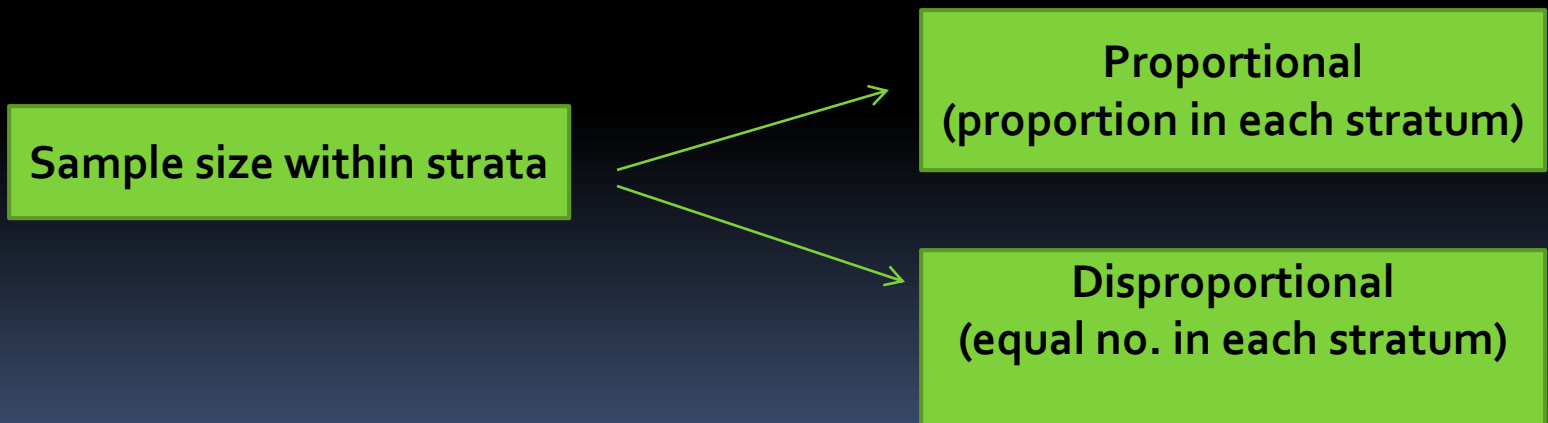


# STRATIFIED RANDOM SAMPLING

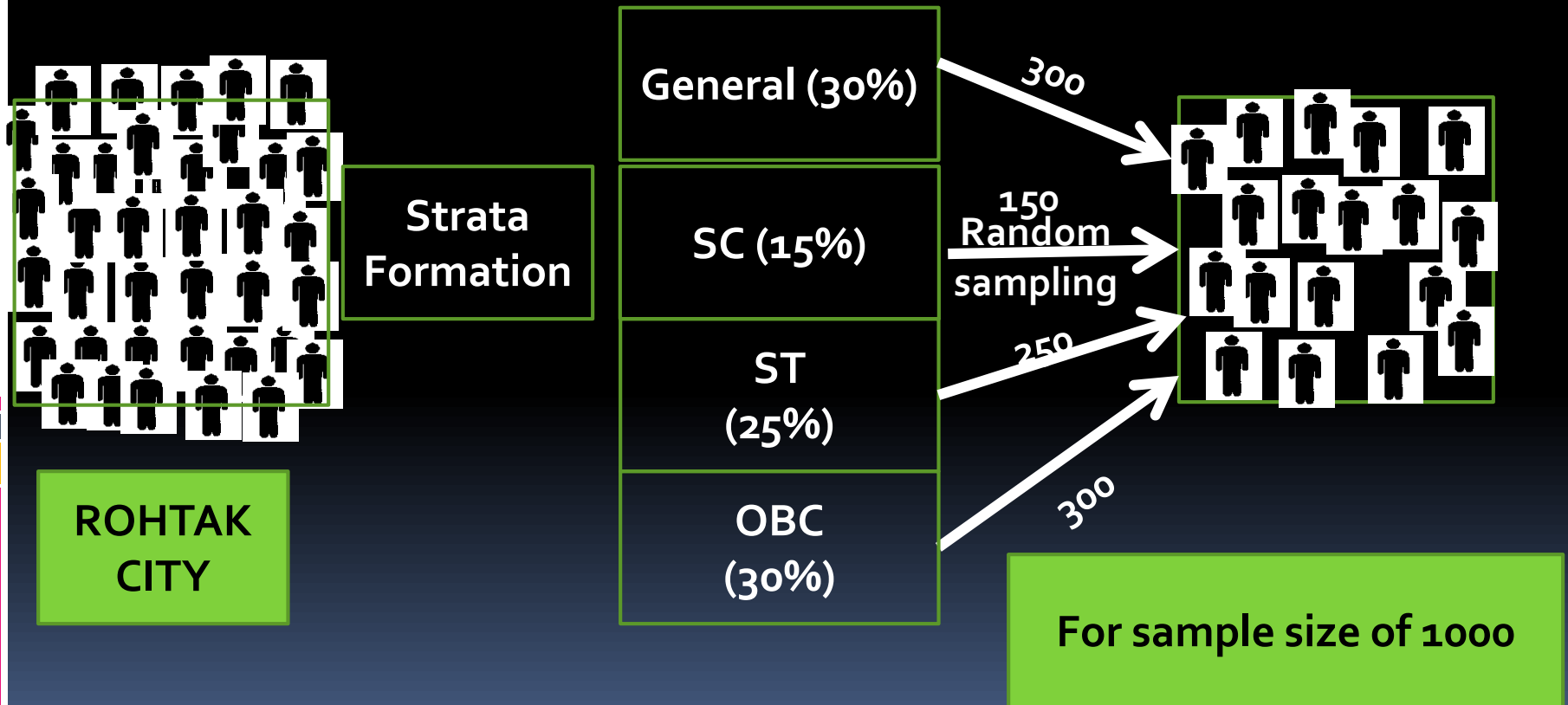
- Universe is sub divided into **mutually exclusive groups**.
  - A **simple random sample** is then chosen independently from each group.
- 

# STRATIFIED RANDOM SAMPLING

- Issues involved in stratification
  - Base of stratification
  - Number of strata
  - Sample size within strata.



# STRATIFIED RANDOM SAMPLING - EXAMPLE



# STRATIFIED RANDOM SAMPLING

- **Merits**

- More representative.
- Greater accuracy.
- Greater geographical concentration.

- **Demerits**

- Utmost care in dividing strata.
- Skilled sampling supervisors.
- Cost per observation may be high.

# SYSTEMATIC SAMPLING

- Selecting **first unit at random.**
- Selecting **additional units at evenly spaced intervals.**
- **Complete list** of population available.

$k=N/n$   
k=sampling interval  
N=universe size  
n=Sample size

Class of 95 students : roll no. 1 to 95  
Sample of 10 students  
k=9.5 or 10  
1<sup>st</sup> student random then every 10<sup>th</sup>

# SYSTEMATIC SAMPLING

- **Merits**

- Simple and convenient.
- Less time consuming.

- **Demerits**

- Population with hidden periodicities.


# CLUSTER SAMPLING

- A sampling technique in which the entire population of interest is divided into **groups, or clusters, and a random sample of these clusters is selected.**
- Each cluster must be **mutually exclusive** and together the clusters must include the entire population .
- After clusters are selected, then all units within the clusters are selected. No units from non-selected clusters are included in the sample.





# CLUSTER SAMPLING

- In cluster sampling, the clusters are the primary sampling unit (PSU's) and the units within the clusters are the secondary sampling units (SSU's)
- 

# STRATIFICATION V/S CLUSTERING

| <b>Stratification</b>  | <b>Clustering</b>   |
|--|---|
| All strata are represented in the sample.                            | Only a subset of clusters are in the sample.              |
| Less error compared to simple random.                                | More error compared to simple random.                     |
| More expensive to obtain stratification information before sampling. | Reduces costs to sample only some areas or Organizations. |

# CLUSTER SAMPLING- STEPS

- Identification of clusters
- List all cities, towns, villages & wards of cities with their population falling in target area under study.
- Calculate cumulative population & divide by 30, this gives sampling interval.
- Select a random no. less than or equal to sampling interval having same no. of digits. This forms 1st cluster.
- Random no.+ sampling interval = population of 2nd cluster.
- Second cluster + sampling interval = 3rd cluster.
- Last or 30th cluster = 29th cluster + sampling interval

# CLUSTER SAMPLING

| • | Freq | c f  | cluster | •      |   |        |                  |
|---|------|------|---------|--------|---|--------|------------------|
| • | I    | 2000 | 2000    | 1      | • | XVI    | 3500 52500 17    |
| • | II   | 3000 | 5000    | 2      | • | XVII   | 4000 56500 18,19 |
| • | III  | 1500 | 6500    |        | • | XVIII  | 4500 61000 20    |
| • | IV   | 4000 | 10500   | 3      | • | XIX    | 4000 65000 21,22 |
| • | V    | 5000 | 15500   | 4, 5   | • | XX     | 4000 69000 23    |
| • | VI   | 2500 | 18000   | 6      | • | XXI    | 2000 71000 24    |
| • | VII  | 2000 | 20000   | 7      | • | XXII   | 2000 73000       |
| • | VIII | 3000 | 23000   | 8      | • | XXIII  | 3000 76000 25    |
| • | IX   | 3500 | 26500   | 9      | • | XXIV   | 3000 79000 26    |
| • | X    | 4500 | 31000   | 10     | • | XXV    | 5000 84000 27,28 |
| • | XI   | 4000 | 35000   | 11, 12 | • | XXVI   | 2000 86000 29    |
| • | XII  | 4000 | 39000   | 13     | • | XXVII  | 1000 87000       |
| • | XIII | 3500 | 44000   | 14,15  | • | XXVIII | 1000 88000       |
| • | XIV  | 2000 | 46000   |        | • | XXIX   | 1000 89000 30    |
| • | XV   | 3000 | 49000   | 16     | • | XXX    | 1000 90000       |

•  $90000/30 = 3000$  sampling interval

# CLUSTER SAMPLING

- **Merits**

- Most economical form of sampling.
- Larger sample for a similar fixed cost.
- Less time for listing and implementation.
- Reduce travel and other administrative costs.

- **Demerits**

- May not reflect the diversity of the community.
- Standard errors of the estimates are high, compared to other sampling designs with same sample size .

# MULTISTAGE SAMPLING

- Sampling process carried out in **various stages**.
- An effective strategy because it banks on multiple randomizations.
- Used frequently when a **complete list of all members of the population does not exist** and is inappropriate.

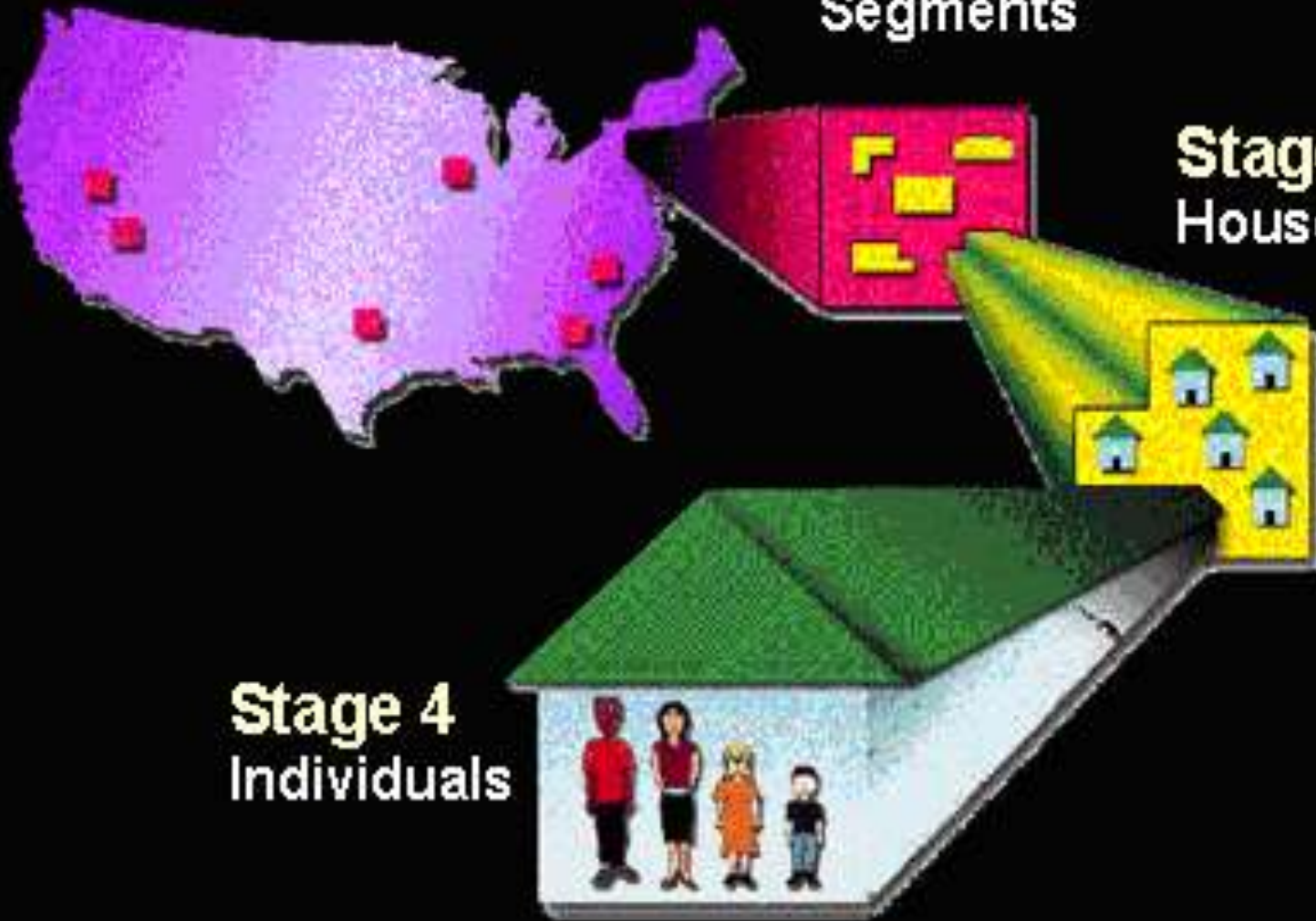
# MULTISTAGE SAMPLING

**Stage 1**  
Counties

**Stage 2**  
Segments

**Stage 3**  
Households

**Stage 4**  
Individuals



# MULTISTAGE SAMPLING

- **Merits**

- Introduces flexibility in the sampling method.
- Enables existing divisions and sub divisions of population to be used as units.
- Large area can be covered.
- Valuable in under developed areas.

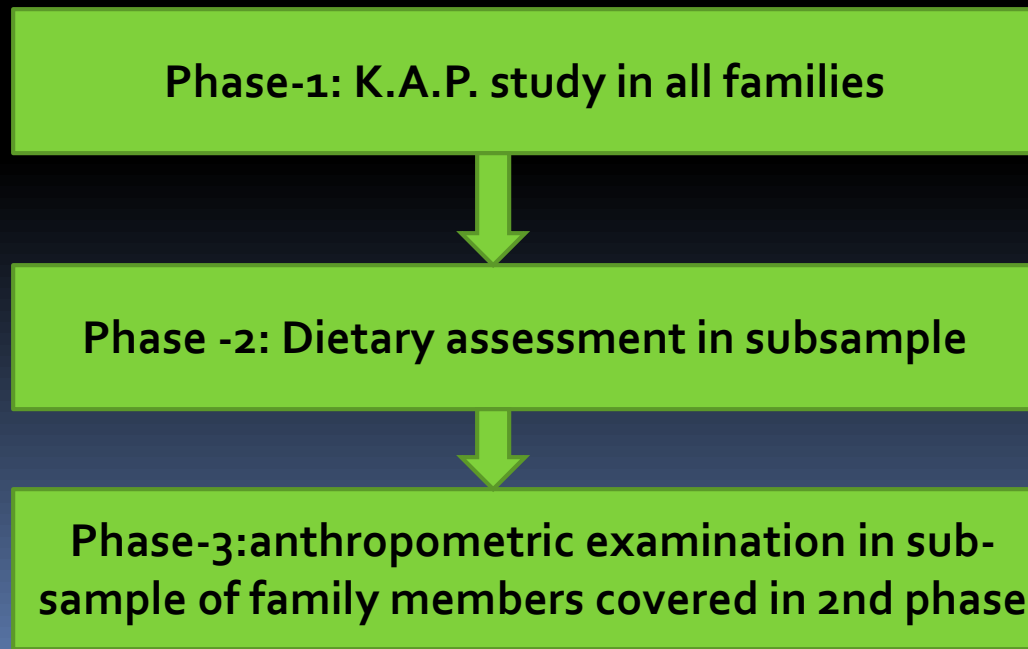
- **Demerits**

- Less accurate than a sample chosen by a single stage process.



# MULTIPHASE SAMPLING

- Used for studies to be carried out in **multiple phases**.
- For e.g. A cross - sectional study on nutrition may be carried out in phases



# LOT QUALITY ASSURANCE SAMPLING

- Originated in the manufacturing industry for quality control purposes.
- Manufacturers were interested in determining whether a batch, or lot, of goods met the desired specifications.
- The only outcome in this type of sampling is “acceptable” or “not acceptable”

# LOT QUALITY ASSURANCE SAMPLING


- The sample size is the number of units that are selected from each lot.
- The decision value is the number of “defective” items that need to be found before the lot is deemed unacceptable.
- There are two types of risks
  - the risk of accepting a “bad” lot, referred to as **Type I error**
  - the risk of not accepting a “good” lot, referred to as **Type II error**.

# LOT QUALITY ASSURANCE SAMPLING

- Information from lots can be combined to obtain the **overall proportion of defects**.
- The population is first divided into a complete set of **non-overlapping lots**.
- Samples are then taken from every lot, and the proportion of defective items in each lot is calculated.
- The LQAS method is an example of stratified sampling, where the lots play the role of the strata.



# LOT QUALITY ASSURANCE SAMPLING

- The advantage of the LQAS method over a traditional stratified sampling design is that the response for each lot is binary (acceptable or not), and therefore smaller sample sizes can be used.
- 

# LOT QUALITY ASSURANCE SAMPLING

- Can be used for evaluating a number of health programmes e.g. immunisation coverage , knowledge of ORS, etc
- Despite successful trials of LQAS in health surveys , its routine use has not been established yet.

# ERRORS

```
graph TD; ERRORS[ERRORS] --> SAMPLING[SAMPLING ERRORS]; ERRORS --> NON_SAMPLING[NON SAMPLING ERRORS]; SAMPLING --> SAMPLE[SAMPLE]; NON_SAMPLING --> SAMPLE_AND_CENSUS[SAMPLE AND CENSUS];
```

SAMPLING  
ERRORS

SAMPLE

NON SAMPLING  
ERRORS

SAMPLE AND  
CENSUS



```
graph TD; A[SAMPLING ERRORS] --> B[BIASED]; A --> C[UNBIASED]; B --> D[ELIMINATION OF ALL SOURCE OF BIAS]; C --> E[INCREASE SAMPLE SIZE]
```

**SAMPLING ERRORS**

**BIASED**

**ELIMINATION OF  
ALL SOURCE OF  
BIAS**

**UNBIASED**

**INCREASE  
SAMPLE SIZE**




# NON SAMPLING ERRORS

- Data specification inadequate & inconsistent with respect to objective of census.
- Inaccurate or inappropriate methods of **interview, observation, definitions.**
- Lack of **trained & experienced investigators.**
- Errors due to **non response.**
- Errors in **data processing operations**
- Errors committed during **presentation.**

**MORE IN COMPLETE ENUMERATION SURVEY**



# REFERENCES

- Methods in Biostatistics by BK Mahajan
  - Statistical Methods by SP Gupta
  - Basic & Clinical Biostatistics by Dawson and Beth.
- 

THANK  
YOU!

