icmr | NIE
INDIAN COUNCIL OF MEDICAL RESEARCH | NATIONAL INSTITUTE OF EPIDEMIOLOGY

Sampling Methods

Dr. R. Ramakrishnan
MAE, PhD

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

Definition of sampling

Procedure by which some members of the population are selected as representatives of the entire population

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

Study population

The study population is the population to which the results of the study will be inferred

The study population depends upon the research question

- How many injections do people received each year in India?
  - Study population: Population of India
- How many needle-sticks health care workers experience each year in India?
  - Study population: Health care workers of India
- How many hospitals have a needle-sticks prevention policy in India
  - Study population: Hospitals of India

The sample needs to be representative of the population in terms of time

- Seasonality
- Day of the week
- Time of the day

The sample needs to be representative of the population in terms of place

- Urban
- Rural

The sample needs to be representative of the population in terms of persons

- Age
- Sex
- Other demographic characteristics

INDIAN COUNCIL OF MEDICAL RESEARCH | NATIONAL INSTITUTE OF EPIDEMIOLOGY

HEALTH RESEARCH FUNDAMENTALS
nie.gov.in

Definition of sampling terms

- Sampling unit (Basic sampling unit, BSU)
  - Elementary unit that will be sampled
    - People
    - Health care workers
    - Hospitals
- Sampling frame
  - List of all sampling units in the population
- Sampling scheme
  - Method used to select sampling units from the sampling frame

Why do we sample populations?

- Obtain information from large populations
- Ensure the efficiency of a study
- Obtain more accurate information

Population

Sample

Practical example

- The Ministry of Health of a country X wants to estimate the proportion of children in elementary schools who have been immunized against childhood infectious diseases
- The task must be completed in one month
- The objective is to estimate the proportion of immunized children

HEALTH RESEARCH FUNDAMENTALS
nie.gov.in

Type of samples

- Non-probability samples
  - Probability of being selected is unknown
  - Convenience samples
    - Biased
    - Best or worst scenario
  - Subjective samples
    - Based on knowledge
    - Time/resources constraints
- Probability samples

Type of samples

- Non-probability samples
- Probability samples
  - Every unit in the population has a known probability of being selected
  - Only sampling method that allows to draw valid conclusions about population

Random sampling in probability samples

- Removes the possibility of bias in selection of subjects
- Ensures that each subject has a known probability of being chosen
- Allows application of statistical theory

Sampling error

- No sample is a perfect mirror image of the population
- Magnitude of error can be measured in probability samples
- Expressed by standard error of mean, proportion, differences…
- Function of:
  - Sample size
  - Variability in measurement

Methods used in probability samples

1. Simple random sampling
2. Systematic sampling
3. Stratified sampling
4. Cluster sampling
5. Multistage sampling

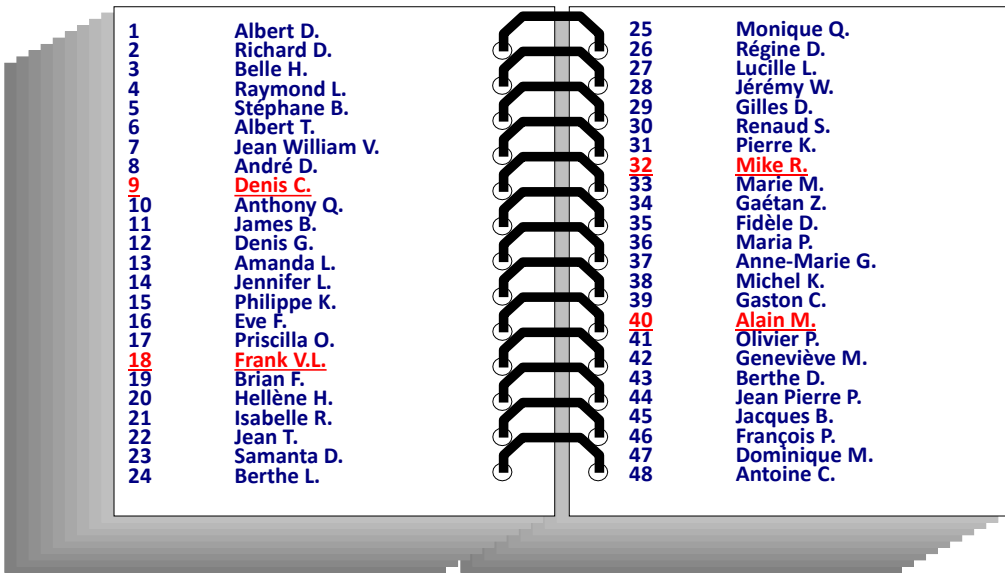
1. Simple random sampling

- Principle
  - Equal chance for each sampling unit
- Procedure
  - Number all units
  - Randomly draw units
- Advantages
  - Simple
  - Sampling error easily measured
- Disadvantages
  - Need complete list of units
  - Does not always achieve best representation

HEALTH RESEARCH FUNDAMENTALS
nie.gov.in
INDIAN COUNCIL OF MEDICAL RESEARCH
NATIONAL INSTITUTE OF EPIDEMIOLOGY

Example of simple random sampling

Numbers are selected at random



HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

icmr NIE INDIAN COUNCIL OF MEDICAL RESEARCH | NATIONAL INSTITUTE OF EPIDEMIOLOGY

2. Systematic sampling

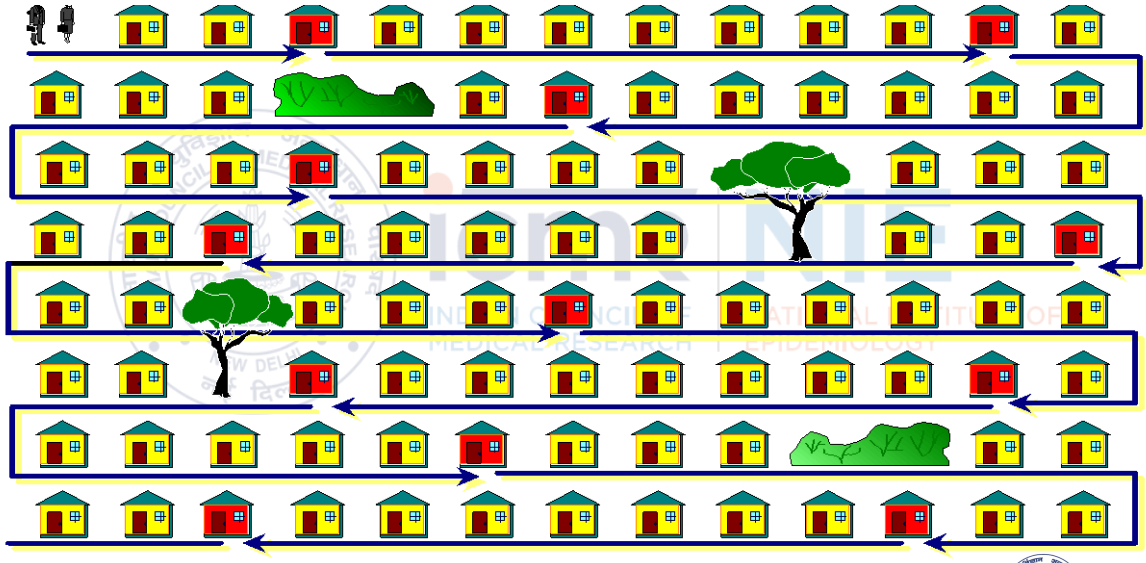
- Principle
  - A unit drawn every k units
  - Equal chance of being drawn for each unit
- Procedure
  - Calculate sampling interval (k = N/n)
  - Draw a random number ($\leq$ k) for starting
  - Draw every k units from first unit
- Advantages
  - Ensures representativity across list
  - Easy to implement
- Disadvantage
  - Dangerous if list has cycles

HEALTH RESEARCH FUNDAMENTALS

nie.gov.in

icmr NIE INDIAN COUNCIL OF MEDICAL RESEARCH | NATIONAL INSTITUTE OF EPIDEMIOLOGY

Example of systematic sampling



Every eighth house is selected

3. Stratified sampling

- Principle
  - Classify population into homogeneous subgroups (strata)
  - Draw sample in each strata
  - Combine results of all strata
- Advantage
  - More precise if variable associated with strata
  - All subgroups represented, allowing separate conclusions about each of them
- Disadvantages
  - Sampling error difficult to measure
  - Loss of precision if small numbers sampled in individual strata

Example of stratified sampling

- Estimate vaccination coverage in a country
- One sample drawn in each region
- Estimates calculated for each stratum
- Each strata weighted to obtain estimate for country

4. Cluster sampling

- Principle
  - Random sample of groups ("clusters") of units
  - All or proportion of units included selected clusters
- Advantages
  - Simple: No list of units required
  - Less travel/resources required
- Disadvantages
  - Imprecise if clusters homogeneous (Large design effect)
  - Sampling error difficult to measure

Cluster sampling

- The sampling unit is not a subject, but a group (cluster) of subjects.
- It is assumed that:
  - The variability among clusters is minimal
  - The variability within each cluster is what is observed in the general population

The two stages of a cluster sample

1. First stage: *Probability proportional to size*
   - Select the number of clusters to be included
   - Compute a cumulative list of the populations in each unit with a grand total
   - Divide the grand total by the number of clusters and obtain the sampling interval
   - Choose a random number and identify the first cluster
   - Add the sampling interval and identify the second cluster
   - By repeating the same procedure, identify all the clusters
2. Second stage
   - In each cluster select a random sample using a sampling frame of subjects (e.g. residents) or households

HEALTH RESEARCH FUNDAMENTALS
nie.gov.in

5. Multistage sampling

- Principle
  - Several chained samples
  - Several statistical units
- Advantages
  - No complete listing of population required
  - Most feasible approach for large populations
- Disadvantages
  - Several sampling lists
  - Sampling error difficult to measure

Key issues

- We cannot study the whole population so we sample it
- Taking a sample leads to sampling error, which is measurable
- Good design and quality assurance ensure validity and while appropriate sample size will ensure precision
- Probability samples are the only one that allow use of statistics as we know them

HEALTH RESEARCH FUNDAMENTALS
nie.gov.in

icmr | NIE
INDIAN COUNCIL OF MEDICAL RESEARCH | NATIONAL INSTITUTE OF EPIDEMIOLOGY

Thank you