روش های نمونه گیری

تهیه و تنظیم:
دکتر نیلوفر معتمد
متخصص پزشکی اجتماعی
عضو هیئت علمی دانشگاه علوم پزشکی بوشهر

Objectives: sampling

To understand:

- Why we use sampling
- Definitions in sampling
- Concept of representativity
- Main methods of sampling
- Sampling errors

Definition of sampling

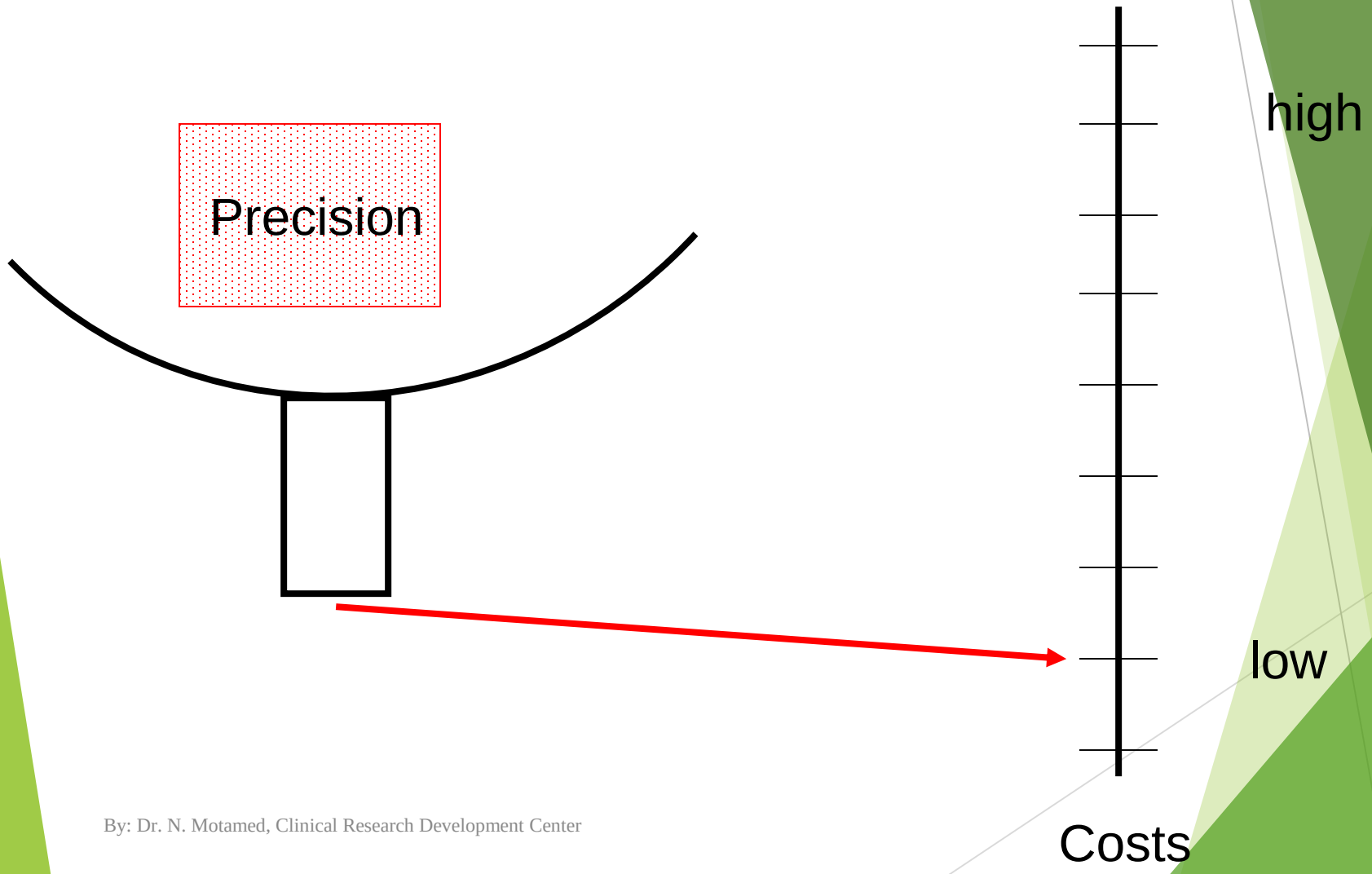
Procedure by which some members of a given population are selected as representatives of the entire population in terms of the desired characteristics

Why bother in the first place?

Get information from large populations with:

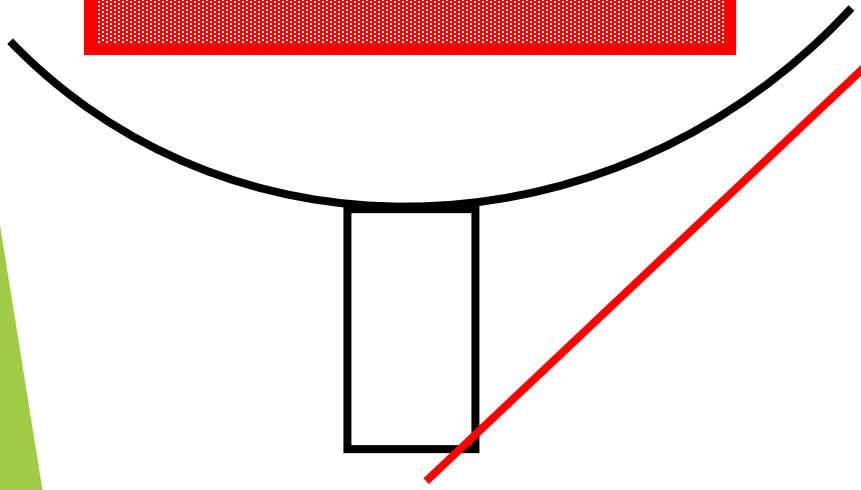
- Reduced costs
- Reduced field time
- Increased accuracy

Sampling



Sampling

Precision



high

low

Costs

Definition of sampling terms

Sampling unit (element)

- Subject under observation on which information is collected
- Example: children <5 years, hospital discharges, health events…

Sampling fraction

- Ratio between sample size and population size
- Example: 100 out of 2000 (5%)

Definition of sampling terms

Sampling frame

- List of all the sampling units from which sample is drawn
  - Lists: e.g. all children < 5 years of age, households, health care units…

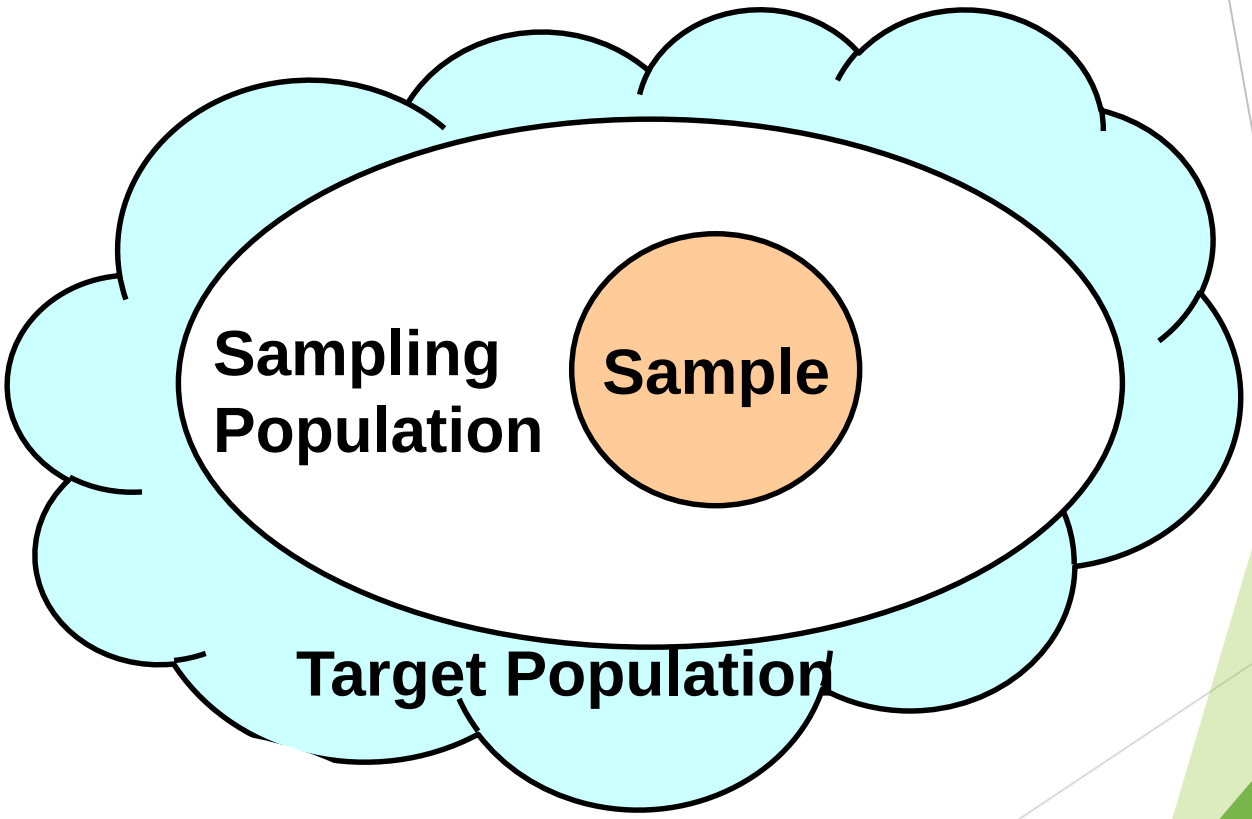
Sampling scheme

- Method of selecting sampling units from sampling frame
  - Randomly, convenience sample…

Survey errors

- Systematic error (or bias)
  - Representativeness (validity)
  - Information bias
- Sampling error (random error)
  - Precision
- Confounding

Sampling and representativeness



Target Population ➔ Sampling Population ➔ Sample

Validity

- Sample should accurately reflect the distribution of relevant variable in population
  - Person (age, sex)
  - Place (urban vs. rural)
  - Time (seasonality)
- Representativeness essential to generalise
- Ensure representativeness before starting
  - Confirm once completed

Information bias

- Systematic problem in collecting information
- Inaccurate measuring
- Scales (weight), ultrasound, lab tests (dubious results)
- Badly asked questions
- Ambiguous, not offering right options…

Sampling error (random error)

- No sample is an exact mirror image of the population
- Standard error depends on
  - size of the sample
  - distribution of character of interest in population
- Size of error
  - can be measured in probability samples
  - standard error

Survey errors: example

Measuring height:

- Measuring tape held differently by different investigators

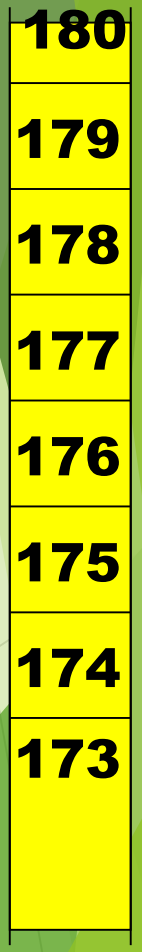
→ loss of precision

→ large standard error

- Tape too short

→ systematic error

→ bias (cannot be corrected afterwards)



Types of sampling

- Non-probability samples
  - Convenience samples
    - Biased
  - Subjective samples
    - Based on knowledge
  - In the presence of time/resource constraints
- Probability samples
  - Random
    - only method that allows valid conclusions about population and measurements of sampling error

Non-probability samples

Convenience samples (ease of access)

Snowball sampling (friend of friend….etc.)

Purposive sampling (judgemental)

You chose who you think should be in the study

Probability of being chosen is unknown

Cheaper- but unable to generalise, potential for bias

Example of a non-probability sample

Take a sample of the population of a Greek island to ask about possible exposures following a gastroenteritis outbreak

Sampling frame: people walking around the port at high noon on a Monday

Probability samples

- Random sampling
  - Each subject has a known probability of being selected
- Allows application of statistical sampling theory to results in order to:
  - Generalise
  - Test hypotheses

Methods used in probability samples

Simple random sampling

Systematic sampling

Stratified sampling

Multi-stage sampling

Cluster sampling

Simple random sampling

- Principle
  - Equal chance/probability of each unit being drawn
- Procedure
  - Take sampling population
    - Need listing of all sampling units (“sampling frame”)
      - Number all units
      - Randomly draw units

Simple random sampling

Advantages

- Simple
- Sampling error easily measured

Disadvantages

- Need complete list of units
- Units may be scattered and poorly accessible
- Heterogeneous population

→ important minorities might not be taken into account

Systematic sampling

Principle

Select sampling units at regular intervals (e.g. every 20th unit)

Procedure

Arrange the units in some kind of sequence

Divide total sampling population by the designated sample size (eg 1200/60=20)

Choose a random starting point (for 20, the starting point will be a random number between 1 and 20)

Select units at regular intervals (in this case, every 20th unit), i.e. 4th, 24th, 44th etc.

Systematic sampling

Advantages

- Ensures representativity across list
- Easy to implement

Disadvantages

- Need complete list of units
- Periodicity-underlying pattern may be a problem (characteristics occurring at regular intervals)

More complex sampling methods

By: Dr. N. Motamed, Clinical Research Development Center

Stratified sampling

When to use

Population with distinct subgroups

Procedure

Divide (stratify) sampling frame into homogeneous subgroups (strata) e.g. age-group, urban/rural areas, regions, occupations

Draw random sample within each stratum

Stratified sampling

Selecting a sample with **probability proportional to size**

Area	Population size	Proportion	Sample size	Sampling fraction
Urban	7000	70%	1000 x 0.7 = **700**	10 %
Rural	3000	30%	1000 x 0.3 = **300**	10 %
Total	**10000**		**1000**	

Stratified sampling

Advantages

- Can acquire information about whole population and individual strata
- Precision increased if variability within strata is smaller (homogenous) than between strata

Disadvantages

- Sampling error is difficult to measure
- Different strata can be difficult to identify
- Loss of precision if small numbers in individual strata (resolved by sampling proportional to stratum population)

Multiple stage sampling

Principle:

- Consecutive sampling
- Example :

sampling unit = household

- 1st stage: draw neighbourhoods
- 2nd stage: draw buildings
- 3rd stage: draw households

Cluster sampling

Principle

- Whole population divided into groups e.g. neighbourhoods
- A type of multi-stage sampling where all units at the lower level are included in the sample
- Random sample taken of these groups (“clusters”)
- Within selected clusters, all units e.g. households included (or random sample of these units)
- Provides logistical advantage

Stage 1: Selection of clusters

- Systematic sampling with probability proportional to size (PPS)
  - Cumulative list of all communities
  - Calculate sampling interval (total pop / 25)
  - Random number identifies first cluster
  - Adding interval to random number identifies subsequent clusters

Village #	Pop.	Cum.
1	309	309
2	169	478
3	359	837
4	182	1019
5	53	1072
6	364	1436
7	435	1871
8	373	2244
9	267	2511
10	380	2891
11	248	3139
12	57	3196
13	513	3709
14	88	3797
15	563	4360
16	261	4621
17	83	4704
18	421	5125
19	187	5312
20	460	5772
...	...	...
97	123	31000

Sampling interval

31000 / 25 = 1240

Random starting point (between 1 and 1240)

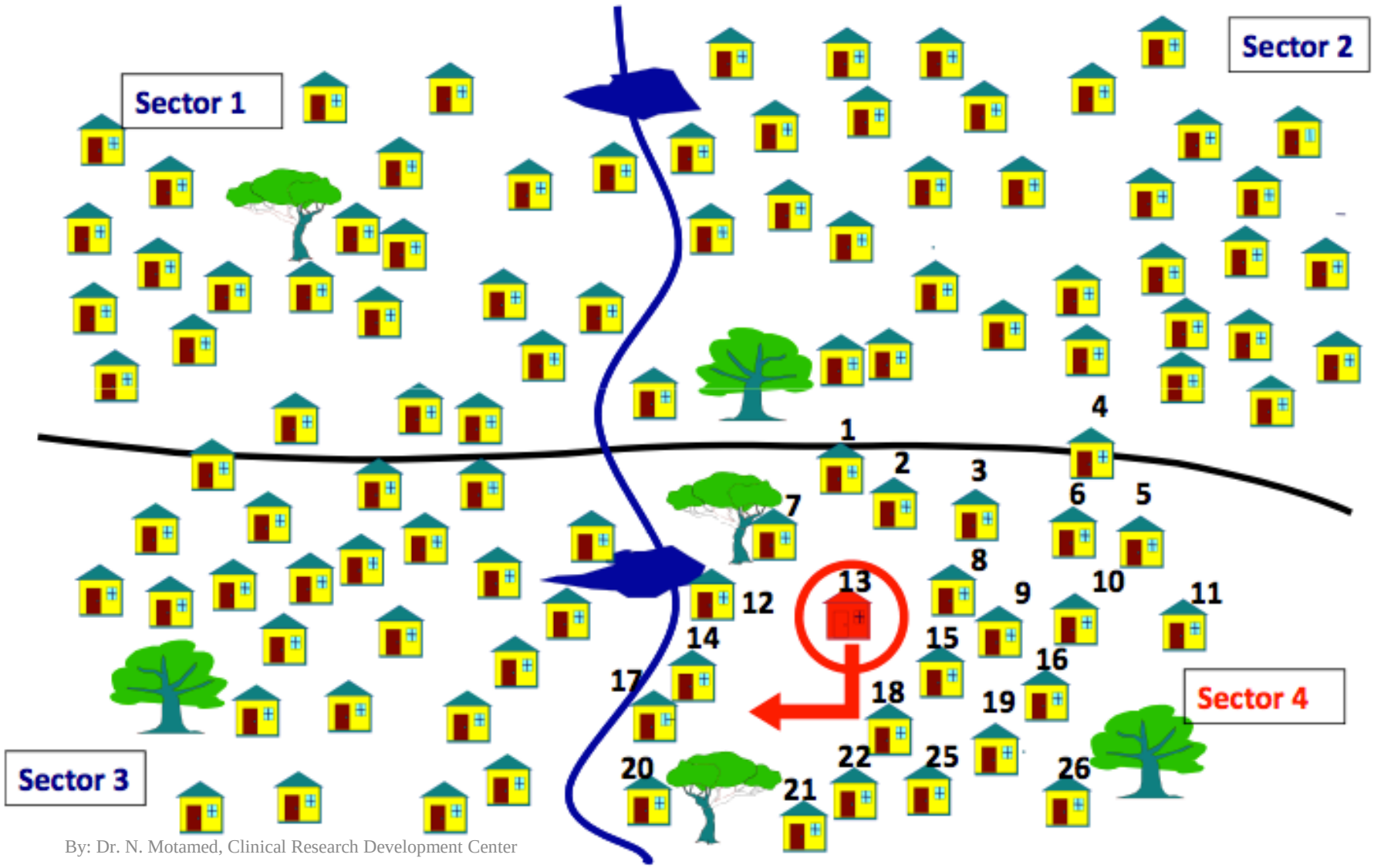
212

1	309	309	212	1	
2	169	478			
3	359	837			
4	182	1019			
5	53	1072			
6	364	1436			+ 1240
7	435	1871	1452	2	
8	373	2244			
9	267	2511			+ 1240
10	380	2891	2692	3	
11	248	3139			
12	57	3196			
13	513	3709			
14	88	3797			+ 1240
15	3563	7360	3932	4,5,6	
16	261	7621			
17	83	7704			
18	421	8125			
19	187	8312			
20	460	8772			
…	…	…			
97	123	31000	31000	25	

Stage 2: Selection of households

- Random selection of first household
  - Divide the village into sectors
  - Select one randomly
  - Number the households
  - Select one randomly
- Move to next household
  - Repeat procedure until 12 individuals are sampled
  - ‘step’
  - One individual per household

Selection of first household in a village

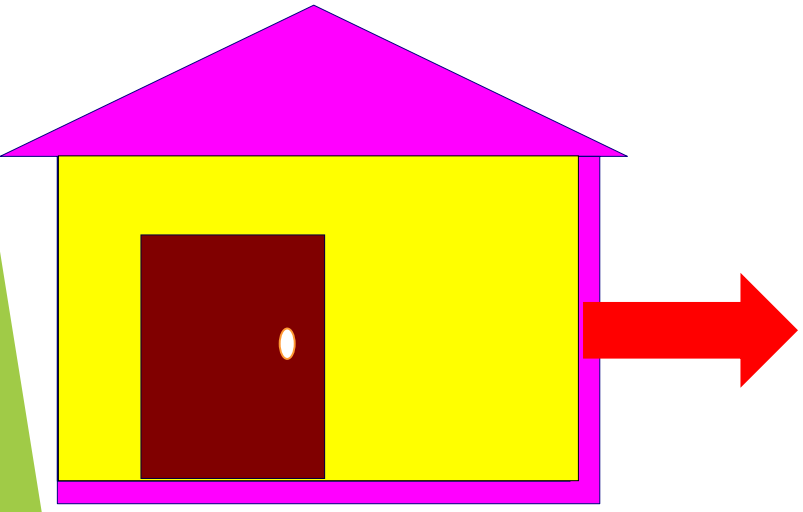


By: Dr. N. Motamed, Clinical Research Development Center

Stage 3: Selection of the sampling unit

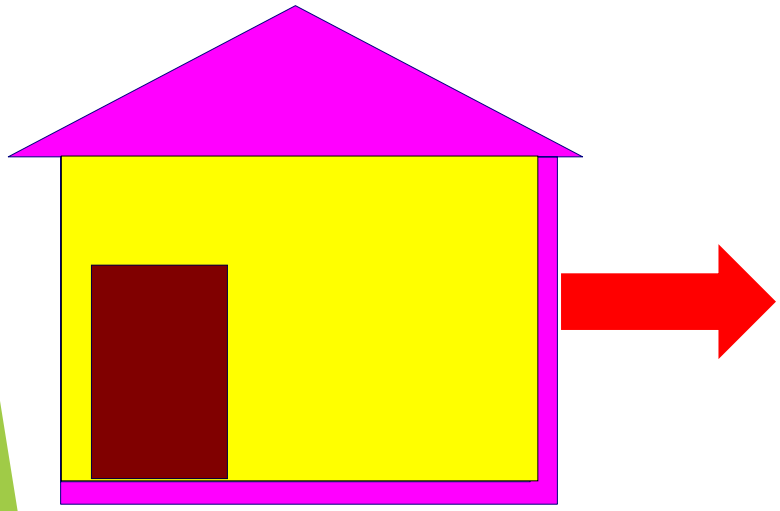
Second-stage units => Households

Third-stage unit => Individuals



Stage 3: Selection of the sampling unit

All third-stage units might be included in the sample



Cluster sampling

Advantages

- Simple as complete list of sampling units within population not required
- Less travel/resources required

Disadvantages

- Cluster members may be more alike than those in another cluster (homogeneous)
- this “dependence” needs to be taken into account in the sample size **and** in the analysis (“design effect”)

Selecting a sampling method

- Population to be studied
  - Size/geographical distribution
  - Heterogeneity with respect to variable
- Availability of list of sampling units
- Level of precision required
- Resources available

Conclusions

- Probability samples are the best
- Ensure
  - Validity
  - Precision
- …..within available constraints