SAMPLING METHODS

Prof Dr Najlaa Fawzi

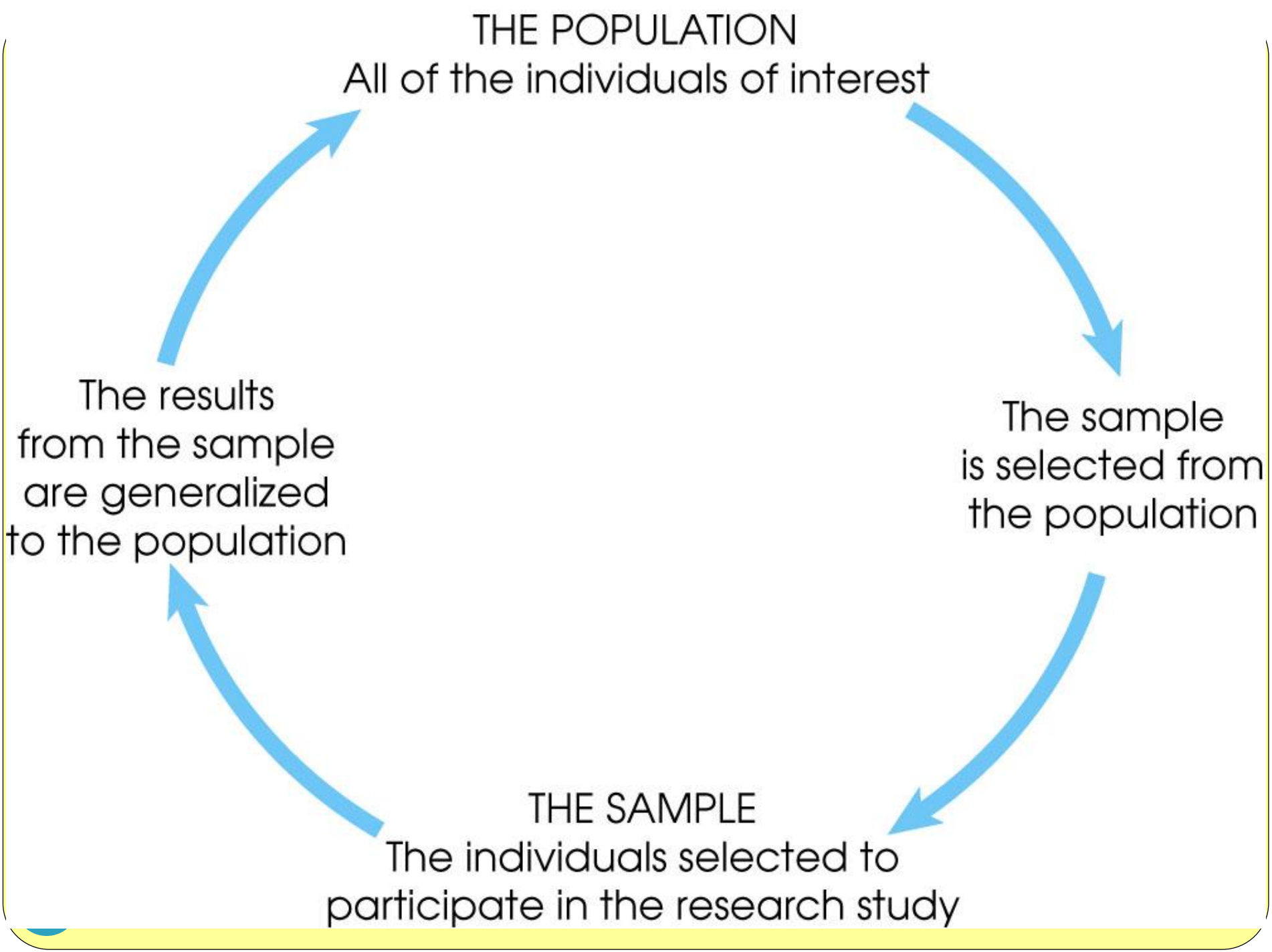
POPULATION:

The largest collection of anything, if this collection has limits, this is finite population, and if not, this is infinite population.

It can be

A-Population of entities: is the largest collection of entities in which we have an interest at a particular time (e.g. population of humans); each population member has many variables.

B-Population of values: it is the largest collection of values of a random variable from which we have an interest of a variable for a particular time e.g. blood urea.


THE POPULATION
All of the individuals of interest
The sample
is selected from
the population
THE SAMPLE
The individuals selected to
participate in the research study
The results
from the sample
are generalized
to the population

SAMPLING

- A sample is "a smaller (but hopefully representative) collection of units from a population used to determine truths about that population" Why sample?
  - Resources (time, money) and workload
  - Gives results with known accuracy that can be calculated mathematically

- What is your population of interest?
  - To whom do you want to generalize your results?
    - All doctors
    - School children
    - Women aged 15-45 years
    - Other
- Can you sample the entire population?

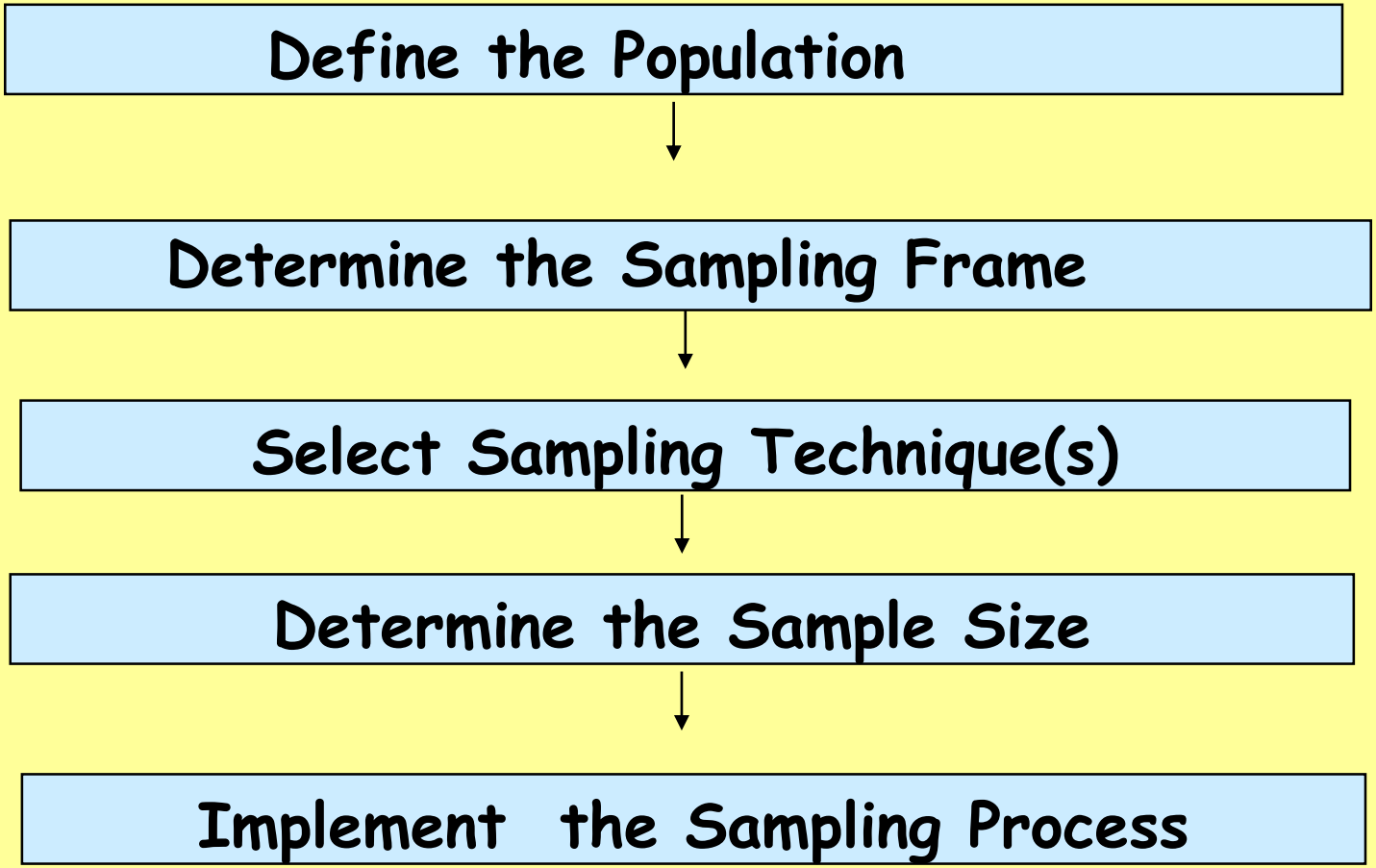
SAMPLING

- 3 factors that influence sample representativeness
  - Sampling procedure
  - Sample size
  - Participation (response)

- When might you sample the entire population?

  - When your population is very small
  - When you have extensive resources
  - When you don't expect a very high response

The Sampling Design Process



Process

- The sampling process comprises several stages:
  - 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 collecting
  - Reviewing the sampling process

Population definition

- 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.

Population definition

- Note also that the population from which the sample is drawn may not be the same as the population about which we actually want information. Often there is large but not complete overlap between these two groups due to frame issues etc.

- Sometimes they may be entirely separate - for instance, we might study rats in order to get a better understanding of human health, or we might study records from people born in 2013 in order to make predictions about people born in 2014.

SAMPLING FRAME

The sampling frame is the list from which the potential respondents are drawn

- Registrar's office
- Class lists
- Must assess sampling frame errors

- *A sampling frame* has the property that we can identify every single element and include any in our sample .

- The sampling frame must be representative of the population

Types of sampling methods: they are probability methods and non-probability methods, the problem in the second type that they cannot be generalized.

Types of Samples

- **Probability (Random) Samples**
- Simple random sample
  - Systematic random sample
  - Stratified random sample
  - Multistage sample
  - Multiphase sample
  - Cluster sample
- **Non-Probability Samples**
  - Convenience sample
  - Purposive sample
  - Quota
  - Snow ball sample

PROBABILITY SAMPLING

- A probability sampling scheme is one in which every unit in the population has a chance (greater than zero) of being selected in the sample, and this probability can be accurately determined.

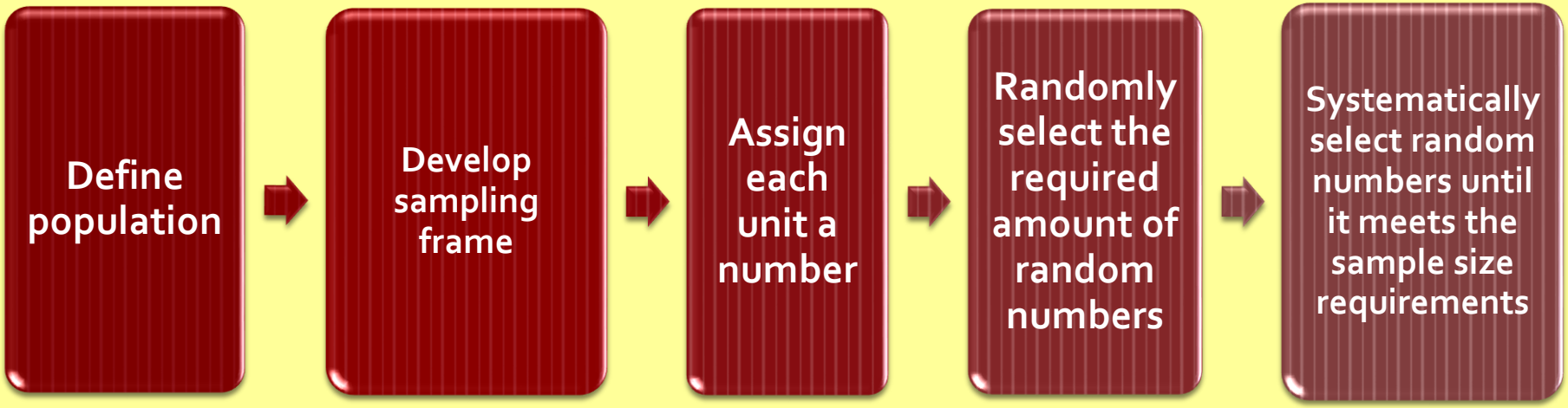
- . When every element in the population *does* have the same probability of selection, this is known as an 'equal probability of selection' (EPS) design. Such designs are also referred to as 'self-weighting' because all sampled units are given the same weight.

PROBABILITY SAMPLING

- Simple Random Sampling,
- Systematic Sampling,
- Stratified Random Sampling,
- Cluster Sampling
- Multistage Sampling.
- Multiphase sampling

SIMPLE RANDOM SAMPLING

Stages in random sampling:



Is also known as '**unrestricted random** sampling'
- Used in clinical trials

SIMPLE RANDOM SAMPLING

- Applicable when population is small, homogeneous & readily available
- All subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection.
- It provides for greatest number of possible samples. This is done by assigning a number to each unit in the sampling frame.
- A table of random number or lottery system is used to determine which units are to be selected.

- Estimates are easy to calculate.
- Simple random sampling is always an EPS design, but not all EPS designs are simple random sampling.

- Disadvantages
- It needs complete list of study population , which is often difficult to obtain.

- If sampling frame large, this method useless.
- Minority subgroups of interest in population may not be present in sample in sufficient numbers for study.

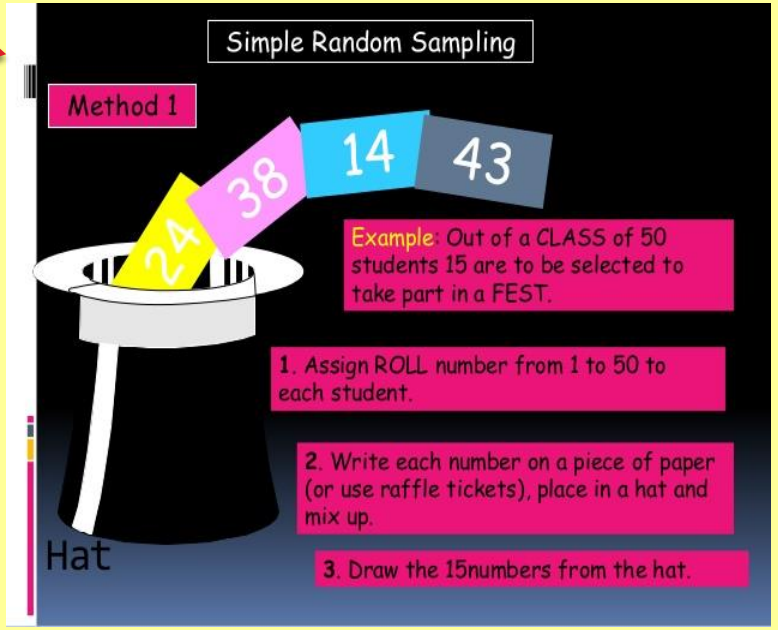
REPLACEMENT OF SELECTED UNITS

- Sampling schemes may be without replacement ('WOR' - no element can be selected more than once in the same sample) or with replacement ('WR' - an element may appear multiple times in the one sample).
- For example, if we catch fish, measure them, and immediately return them to the water before continuing with the sample, this is a WR design, because we might end up catching and measuring the same fish more than once. However, if we do not return the fish to the water (e.g. if we eat the fish), this becomes a WOR design.

Methods of Simple random sampling:

- Lottery method
- Random no. tables

* Computer software.



SYSTEMATIC SAMPLING

- Systematic sampling depend on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list.

- This method is preferred when the population is large , scattered and not homogenous .

Systematic sampling involves a random start and then proceeds with the selection of every *k*th element from then onwards.

In this case, *k*=(population size/sample size). It is important that the starting point is not automatically the first in the list, but is instead randomly chosen from within the first to the *k*th element in the list.

Systematic Random Sampling

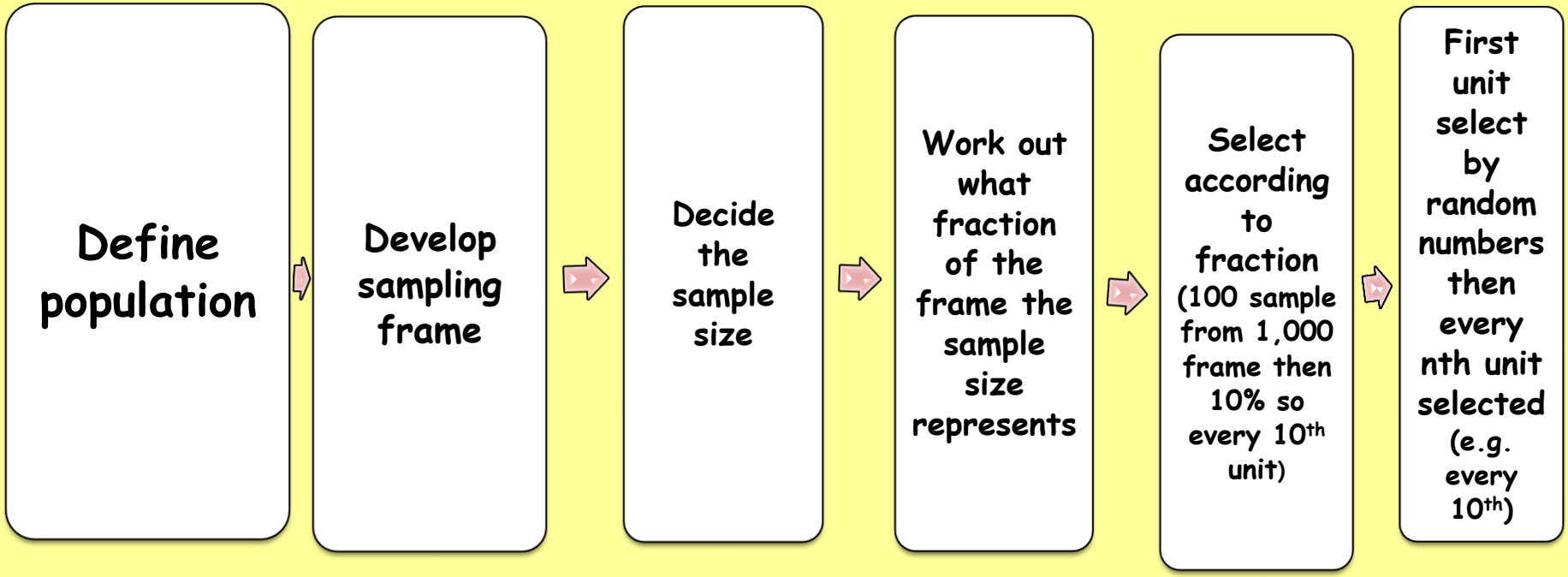
-- Based on sampling fraction: Every Kth unit is chosen in the population list,
where K is chosen by sampling interval

-- Sampling Interval (K) Q = Total no. of units in population/ Total no. of units in
sample

-- Applicable for large, non-homogenous populations where complete list of individuals is available

-- For example, if there is a population of 1000 from which sample of 20 is to be
chosen, then K = 1000/20 = 50; thus every 50th unit will be included in the
sample (i.e. 1st, 51st, 101st, so on…)
First unit among first 50 is chosen by simple random sampling.

SYSTEMATIC SAMPLING……



As described above, systematic sampling is an EPS method, because all elements have the same probability of selection (in the example given, one in ten). It is *not* 'simple random sampling' because different subsets of the same size have different selection probabilities - e.g. the set {4,14,24,...,994} has a one-in-ten probability of selection, but the set {4,13,24,34,...} has zero probability of selection.

- **ADVANTAGES:**
- Sample easy to select
- Suitable sampling frame can be identified easily
- Sample evenly spread over entire reference population
- Time and labour for sample collection is relatively small.
- **DISADVANTAGES:**
- Sample may be biased if hidden periodicity in population coincides with that of selection.
- Difficult to assess precision of estimate from one survey.

STRATIFIED SAMPLING

This method used when the population is not homogenous and is composed of diverse segments.

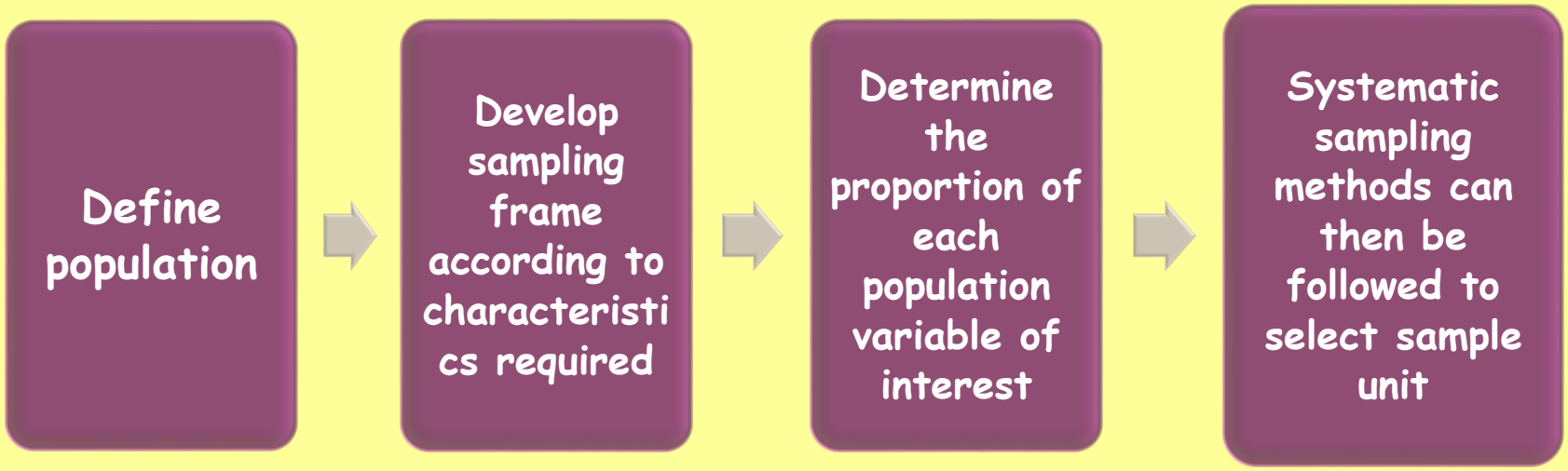
Where population embraces a number of distinct categories, the frame can be organized into separate "strata." Each stratum is then sampled as an independent sub-population, out of which individual elements can be randomly selected.

Every unit in a stratum has same chance of being selected.

Every unit in a stratum has same chance of being selected. This method gives more representative sampling than simple random sampling in a given large population.

Stratified Random Sampling

-Non-homogenous population is converted to homogenous groups/classes
(strata); sample is drawn from each strata at random, in proportion to its size
-Applicable for large non-homogenous population
-- Gives more representative sample than simple random sampling
-None of the categories is under or over-represented
-- For example, In a population of 1000, sample of 100 is to be drawn for
Hemoglobin estimation; first convert non-homogenous population is converted
to homogenous strata (i.e. 700 males and 300 females), then draw 70 males and
30 females randomly respectively


Define population
Develop sampling frame according to characteristics required
Determine the proportion of each population variable of interest
Systematic sampling methods can then be followed to select sample unit

Using same sampling fraction for all strata ensures proportionate representation in the sample.

Adequate representation of minority subgroups of interest can be ensured by stratification & varying sampling fraction between strata as required.

- Finally, since each stratum is treated as an independent population, different sampling approaches can be applied to different strata.

- Types of Stratified Samples
  - Proportional Stratified Sample:
    - The number of sampling units drawn from each stratum is in proportion to the relative population size of that stratum
  - Disproportional Stratified Sample:
    - The number of sampling units drawn from each stratum is allocated according to analytical considerations e.g. as variability increases sample size of stratum should increase

- **Optimal allocation stratified sample**
  - The number of sampling units drawn from each stratum is determined on the basis of both size and variation.
  - Calculated statistically

Advantage

- It is more representative
- It gives estimates with increased precision
- As the population is more concentrated, the time and money will be saved.

Disadvantage

 Requires accurate information on proportions of each stratum

It is very difficult task to divide the population into homogenous strata .

 Stratified lists costly to prepare. This may require considerable time , money and statistical expertise.

Disadvantages to using stratified sampling.
First, sampling frame of entire population has to be prepared separately for each stratum
Second, when examining multiple criteria, stratifying variables may be related to some, but not to others, further complicating the design, and potentially reducing the utility of the strata.

Finally, in some cases (such as designs with a large number of strata, or those with a specified minimum sample size per group), stratified sampling can potentially require a larger sample than would other methods

Select a stratified random sample of 20 patients from 200 patients.

	Disease A	Disease B	Disease C	Disease D	TOTAL
No. of patients	100	60	20	20	200
%	50	30	10	10	100

Out of 20 patients the no. to be selected wise area.

Strata Disease A (100/ 200)X 20 = 50% OF 20 = 10

Disease B(60/ 200)X 20 = 30% OF 20 =6

Disease C(20/ 200)X 20 = 10% OF 20 =2

Disease D(20/ 200)X 20 = 10% OF 20 =2

Cluster Random Sampling

-Applicable when units of population are natural groups or clusters.

* Clusters are heterogeneous within themselves but homogenous with respect to each other

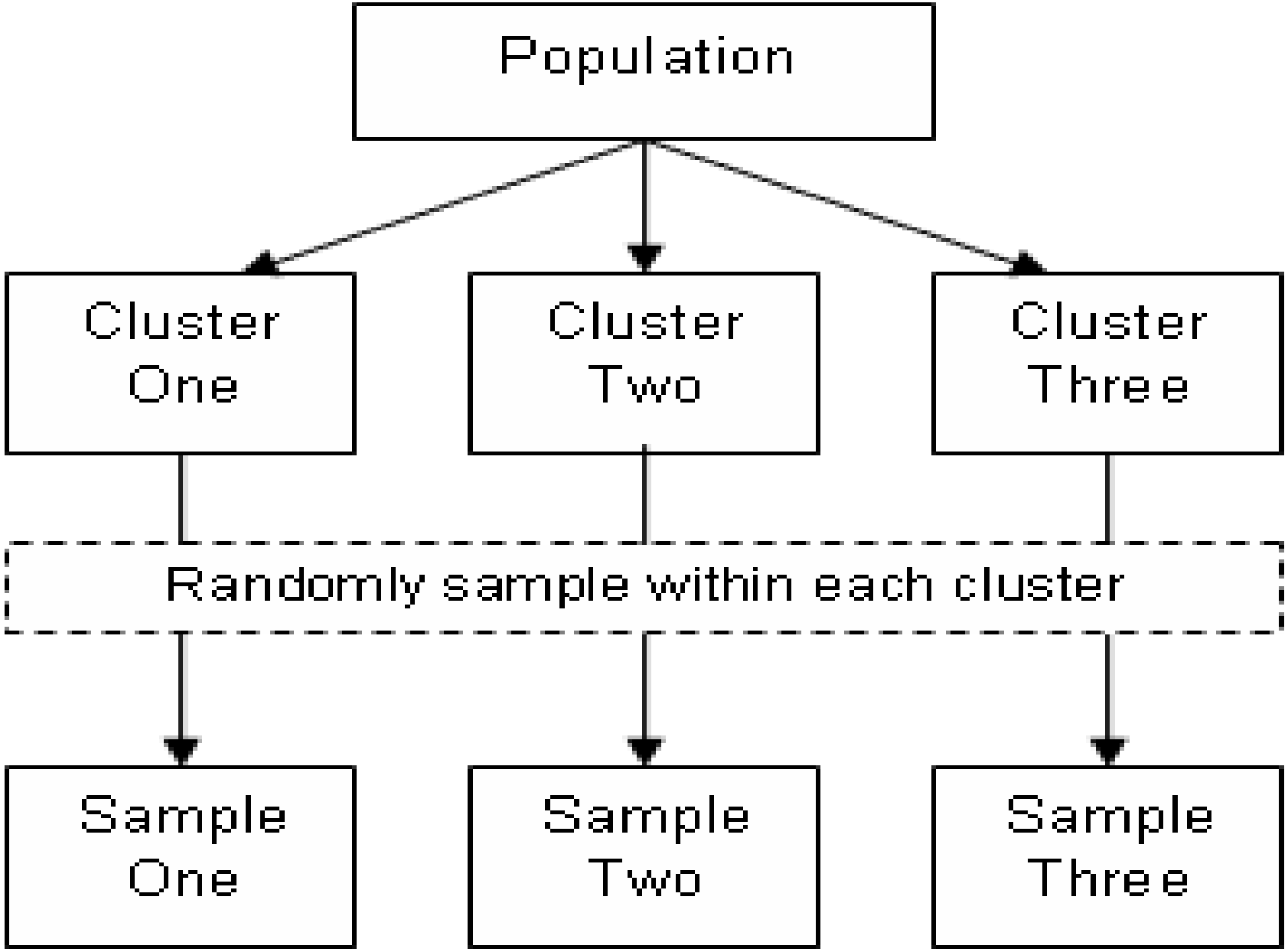
- Often used to evaluate vaccination coverage in EPI

- Cluster sampling is an example of 'two-stage sampling' .
- First stage a sample of areas is chosen;
- Second stage a sample of respondents *within* those areas is selected.
- Population divided into clusters of homogeneous units, usually based on geographical contiguity.
- Sampling units are groups rather than individuals.
- A sample of such clusters is then selected.
- All units from the selected clusters are studied.

Two types of cluster sampling methods

One-stage sampling. All of the elements within selected clusters are included in the sample.

Two-stage sampling. A subset of elements within selected clusters are randomly selected for inclusion in the sample.


Population
Cluster One
Cluster Two
Cluster Three
Randomly sample within each cluster
Sample One
Sample Two
Sample Three

- **Advantages :**
- Cuts down on the cost of preparing a sampling frame.
- This can reduce travel and other administrative costs.
- Requires list of all clusters, but only of individuals within chosen clusters
- Can estimate characteristics of both cluster and population

- Sampling interval is also calculated in CRS
- Accuracy : Low error rate of only ± 5%

Disadvantages:

Sampling error is higher for a simple random sample of same size. Clusters cannot be compared with each other

Often used to evaluate vaccination coverage in EPI

Use in India: Evaluation of immunization coverage

-- WHO technique used: 30 × 7 technique (total = 210 children)

-- WHO technique used in CRS: 30 × 7 technique (total = 210 children)

* 30 clusters, each containing
* 7 children who are 12 – 23 months age and are completely immunized for primary immunization (till Measles vaccine)

Difference Between Strata and Clusters

- Although strata and clusters are both non-overlapping subsets of the population, they differ in several ways.
-
- All strata are represented in the sample; but only a subset of clusters are in the sample.

- With stratified sampling, the best survey results occur when elements within strata are internally homogeneous.
- However, with cluster sampling, the best results occur when elements within clusters are internally heterogeneous

Moreover, by avoiding the use of all sample units in all selected clusters, multistage sampling avoids the large, and perhaps unnecessary, costs associated with traditional cluster sampling.

MULTISTAGE SAMPLING

- Complex form of cluster sampling in which two or more levels of units are embedded one in the other.
- First stage, random number of districts chosen in all states.
- Followed by random number of towns , villages.
- Then third stage units will be houses.
- All ultimate units (houses, for instance) selected at last step are surveyed.

- This technique, is essentially the process of taking random samples of preceding random samples.

- Not as effective as true random sampling, but probably solves more of the problems inherent to random sampling.

-

- An effective strategy because it banks on multiple randomizations. As such, extremely useful.

- Multistage sampling used frequently when a complete list of all members of the population not exists and is inappropriate.

Multistage Random Sampling

-Is done in successive stages; each successive sampling unit is nested in the previous sampling unit. For example, in large country surveys, states are chosen, then districts, then villages, then every 10th person in village as final sampling unit.

Advantage:

- Introduces flexibility in sampling. This method is very helpful in many large scale surveys where population list preparation is difficult.

- It is less expensive and less time consuming.
- It permits available resources to be concentrated on limited numbers of units of the frame.

Disadvantages

- Sampling error is usually increased.
- Sampling units will be of unequal size at various stages resulting in analytical difficulties.

MULTI PHASE SAMPLING

- Is done in successive phases
- Part of the information collected from whole sample & part from subsample.

- Study of nutrition ,all the families in the original sample are covered for KAP study in 1st phase. A sub –sample of the families is than surveyed for dietary intake in 2nd phase
- Then a sub-sample of family members covered in 2nd phase is subjected to anthropometric examination in 3rd phase.
- Survey by such procedure is less costly, less difficult & more purposeful

A further sub –sample from 3rd phase is subjected to Biochemical tests in 4th phase
Thus the number of subject or units gets reduced in every Succeeding phase, thereby reducing the magnitude of the Complicated and costly procedure reserved for the last Phase.

Survey by such procedure is less costly, less difficult & more purposeful

MATCHED RANDOM SAMPLING

A method of assigning participants to groups in which pairs of participants are first matched on some characteristic and then individually assigned randomly to groups.

- The Procedure for Matched random sampling can be briefed with the following contexts.

- Two samples in which the members are clearly paired, or are matched explicitly by the researcher. For example, IQ measurements or pairs of identical twins.

- Those samples in which the same attribute, or variable, is measured twice on each subject, under different circumstances. Commonly called repeated measures.

- Examples include the times of a group of athletes for 1500m before and after a week of special training; the milk yields of cows before and after being fed a particular diet.

NONPROBABILITY SAMPLING

- Any sampling method where some elements of population have no chance of selection (these are sometimes referred to as 'out of coverage'/'under covered'), or where the probability of selection can't be accurately determined.

- It involves the selection of elements based on assumptions regarding the population of interest, which forms the criteria for selection.

-

- Hence, because the selection of elements is nonrandom, nonprobability sampling not allows the estimation of sampling errors..

Non random sampling (non-probability sampling) technique is used when it cannot be ensured that each item has an equal chance of being selected , or when selection is based on expert knowledge of the population.

- Nonprobability Sampling includes: Accidental Sampling, Quota Sampling and Purposive Sampling.

- In addition, nonresponse effects may turn *any* probability design into a nonprobability design if the characteristics of nonresponse are not well understood, since nonresponse effectively modifies each element's probability of being sampled.

02. Non Probability Samples

A non probability sample relies on the researcher selecting the respondents.

They are considered to be:

- Interpretive
- Subjective
- Not scientific
- Qualitative
- Unrepresentative

The purpose of this method is to make an clear choice on researcher own judgment about exactly whom to include in the sample.

Despite various limitations and criticisms, there are numerous advantages of non-probability methods:

- Cheaper
- Used when sampling frame is not available

- Useful when population is so widely dispersed that cluster sampling would not be efficient.
- Often used in exploratory studies, e.g. for hypothesis generation
- Some research not interested in working out what proportion of population gives a particular response but rather in obtaining an idea of the range of responses on ideas that people have.

QUOTA SAMPLING

- The population is first segmented into mutually exclusive sub-groups, just as in stratified sampling.
-
- Then judgment used to select subjects or units from each segment based on a specified proportion.

- For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60.
- It is this second step which makes the technique one of non-probability sampling.
- In quota sampling the selection of the sample is non-random.

For example interviewers might be tempted to interview those who look most helpful.

The problem is that these samples may be biased because not everyone gets a chance of selection.

This random element is its greatest weakness and quota versus probability has been a matter of controversy for many years

Stratified versus Quota Sample

Similarities:

- Population is divided into segments (strata).
- Elements are selected from each segment.

Key Difference:

- Stratified sampling uses probability methods.
- Quota samples are based on a researcher's judgment.
- Therefore, stratified sampling allows the establishment of the sampling distribution, confidence intervals and statistical tests.

CONVENIENCE SAMPLING

- Sometimes known as grab or opportunity sampling or accidental or haphazard sampling.
- A type of non probability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, readily available and convenient.
- The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough.

This is termed a convenience sample, it has obvious advantage in cost and logistics, and it is a good choice for many research questions .

A convenience sample can minimize volunteerism and other selection biases by consecutively selecting every accessible person who meets the entry criteria .

Convenience Sampling

-- Patients are selected, in part or in whole, at the convenience of the researcher;
no/limited attempt to ensure that sample is an accurate representation of Population

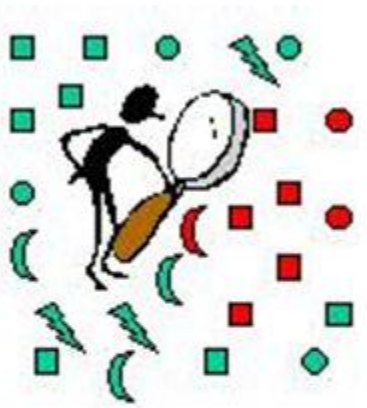
-- For example, standing at a shopping mall and selecting shoppers as they walk by to fill out a survey

Non-probability Sampling Methods

Convenience sampling method

the selecting on the basis of convenience

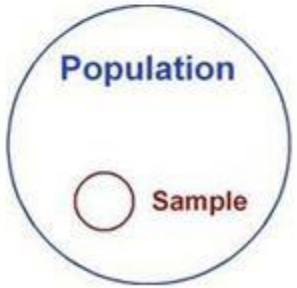
the selection at familiar locations and to choose respondents who are like themselves



often used during preliminary research efforts to get a gross estimate of the results

Judgment method

selecting samples that require a judgment or an "educated guess"



must be confident that the chosen sample is truly representative of the entire population.

A consecutive sample is especially desirable when it amounts to taking the entire accessible population over along enough period to include the seasonal variations or other changes over time considered important to the study question.

This sampling technique is also useful in documenting that a particular quality of a substance or phenomenon occurs within a given sample. Such studies are also very useful for detecting relationships among different phenomena.

Criticisms

The most obvious criticism about convenience sampling is sampling bias and that the sample is not representative of the entire population.

Another significant criticism about using a convenience sample is the limitation in generalization and inference making about the entire population.

For example, if the interviewer was to conduct a survey at a shopping center early in the morning on a given day, the people that he/she could interview would be limited to those given there at that given time, which would not represent the views of other members of society in such an area, if the survey was to be conducted at different times of day and several times per week.

This type of sampling is most useful for pilot testing.

In social science research, snowball sampling is a similar technique, where existing study subjects are used to recruit more subjects into the sample.

This sampling technique is often used in hidden population which are difficult for researchers to access.

Judgmental sampling or Purposive sampling

- - The researcher chooses the sample based on who they think would be appropriate for the study.

- This is used primarily when there is a limited number of people that have expertise in the area being researched

When selecting a sampling strategy it is necessary that it fits the purpose of the study, the resources available, the question being asked and the limitations being faced.

This holds true for sampling strategy as well as sample size.

A sample is expected to mirror the population from which it comes, however, there is no guarantee that any sample will be precisely representative of the population which it comes.

In practice, it is rarely known when a sample is unrepresentative and should be discarded.

ERRORS IN SAMPLING

What can make a sample unrepresentative of its population?

One of the most frequent causes is

sampling error

Sampling error comprises the differences between the sample and the population that are due totally to the particular units that happen to have been selected.

Error is the difference between observed value and true value.

Sampling error

- The statistics of different samples from same population: different each other!
- The statistics: different from the parameter!

For example, suppose that a sample of 100 women are measured and are all found to be taller than178cm.

It is very clear even without any statistical prove that this would be a highly unrepresentative to invalid conclusions.

The sampling error exists in any sampling research.

It can not be avoided but may be estimated.

Three kinds of error:

(1) Systematic error (fixed)

(2) Measurement error (random)

(Observational error)

(3) Sampling error (random)

Errors in Sampling

There are two basic causes for sampling error:

1-Sampling bias is a tendency to favor the units that have particular characteristics.

Sampling bias is usually the result of a poor sampling plan.

BIAS IN SAMPLING-5 sources

- Any deviation from rules-
  self selection → volunteers
- Elimination of hard to identify people
  missing persistent absentees
- Replacement of previously selected individuals
  - Difficult to trace after being included in frame/uncooperative
- Large scale refusal
- List/sampling frame goes out of date

2-chance : That is the error that occurs just because of bad luck.

This may result in untypical choices.

Unusual units in a population do exist and there is always a possibility that an abnormally large number of them will be chosen.

The main protection against this kind of error is to use a large enough sample.

FACTORS THAT INFLUENCING THE SAMPLING ERROR ARE:

1- The size of the sample

2- The natural variability of the individual readings

As the size of the sample increases , sampling error will decrease.

As the individual reading vary widely from one another, we get more variability from one sample to another.

Errors of Observation

- Interview error- interaction between interviewer and person being surveyed
- Respondent error: respondents have difficult time answering the question
- Measurement error: inaccurate responses when person doesn't understand question or poorly worded question
- Errors in data collection

Non-Observation Errors

- Sampling error: naturally occurs
- Coverage error: people sampled do not match the population of interest
- Underrepresentation
- Non-response: won't or can't participate

Non sampling error (measurement error)

The other main cause of unrepresentative samples is non sampling error.

Like sampling error, non sampling error may either be produced by participants in the statistical study or be an innocent by product of the sampling procedures.

These are often more important than the sampling errors.

A non sampling error is an error that results from the manner in which the observations are made.

Induced bias :Finally, it should be noted that the personal preconceptions of either the designer of the study or the data collector may tend to induce bias.

To protect against induced bias, advice of an individual trained in statistics should be required in the design and someone else aware of search pitfalls should serve in an auditing capacity.

What sampling method u recommend?

- Determining proportion of undernourished five year olds in a village.
- Investigating nutritional status of preschool children.
- Selecting maternity records for the study of previous abortions or duration of postnatal stay.
- In estimation of immunization coverage in a province, data on seven children aged 12-23 months in 30 clusters are used to determine proportion of fully immunized children in the province.
- Give reasons why cluster sampling is used in this survey.

Practice

A large elementary school has 15 classrooms, with 24 children in each classroom. A sample of 30 children is chosen by the following procedure:

Each of the 15 teachers selects 2 children from his or her classroom to be in the sample by numbering the children from 1 to 24, using a random digit table to select two different random numbers between 01 and 24. The 2 children with those numbers are in the sample.

Did this procedure give a simple random sample of 30 children from the elementary school?

a) No, because the teachers were not selected randomly

b) No, because not all possible groups of 30 children had the same chance of being chosen

c) No, because not all children had the same chance of being chosen

d) Yes, because each child had the same chance of being chosen

e) Yes, because the numbers were assigned randomly to the children