Sampling

Population

A population can be defined as including all people or items with the characteristic one wishes to understand.

To whom do you want to generalize your results?
Examples:

- All doctors
- School children
- Indians
- Women aged 15-45 years

Sample

- A sample is “a smaller (but hopefully representative) collection of units from a population used to determine truths about that population” (Field, 2005)

- Why sample?
  - Resources (time, money) and workload
  - Gives results with known accuracy that can be calculated mathematically

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

Population and Sample

- 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 records from people born in 2008 in order to make predictions about people born in 2009.

Census Sample

- A census study occurs if the entire population is very small or it is reasonable to include the entire population (for other reasons).

- It is called a census sample because data is gathered on every member of the population.

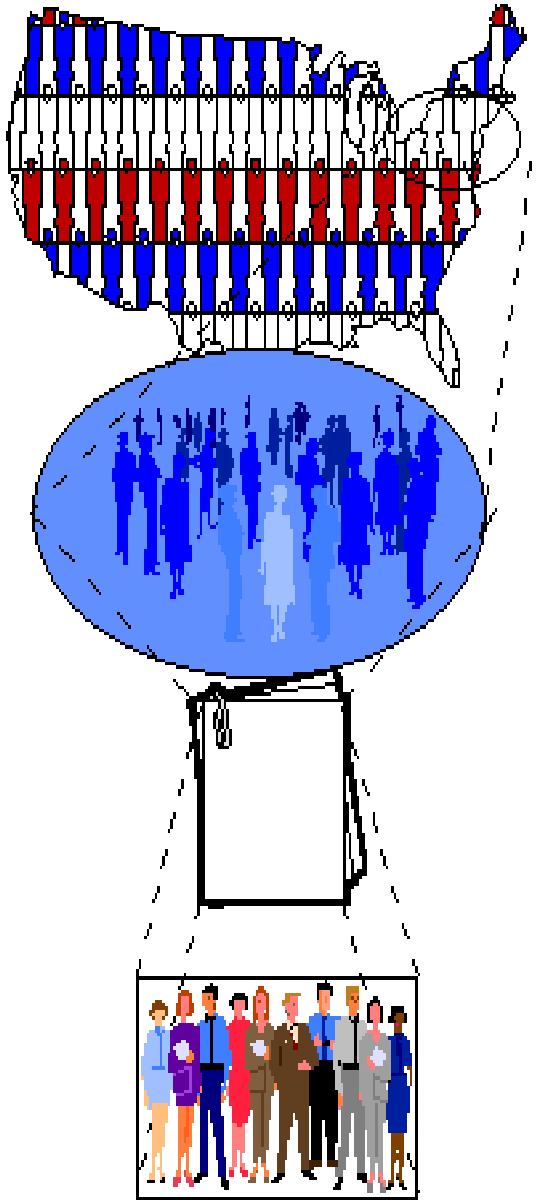
SAMPLING…….

- 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

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

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?

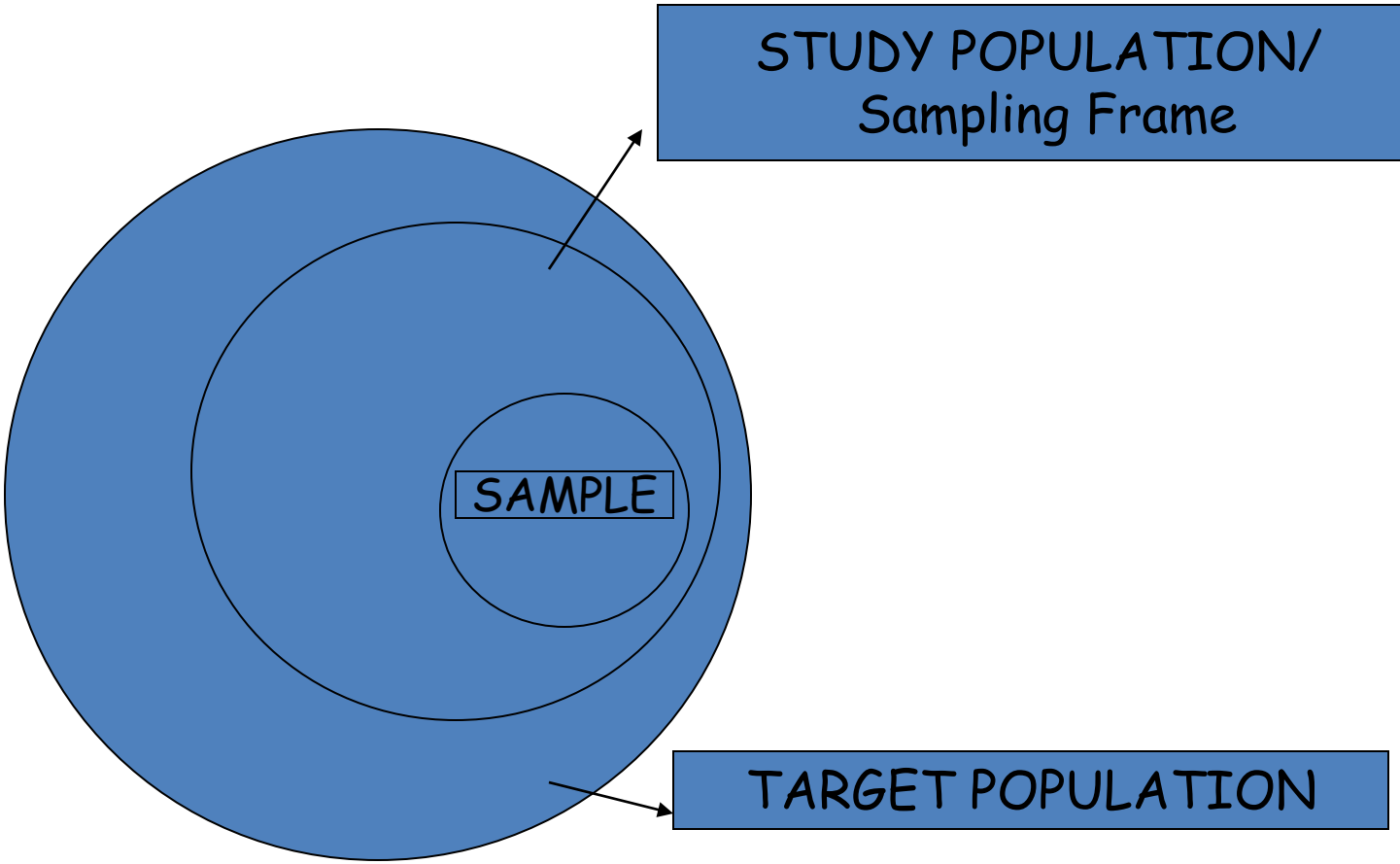
The Sampling Frame

Who is in your study?

The Sample

SAMPLING BREAKDOWN

SAMPLING.....


STUDY POPULATION/
Sampling Frame
SAMPLE
TARGET POPULATION

Target Population:

The population to be studied/ to which the investigator wants to generalize his results

Sampling Unit:

smallest unit from which sample can be selected

Sampling frame

List of all the sampling units from which sample is drawn

Sampling scheme

Method of selecting sampling units from sampling frame

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 scheme for selecting items or events from the frame
  - Determining the sample size
  - Implementing the sampling plan
  - Sampling and data collecting

Sampling Frame

- In statistics, a sampling frame is the source material or device from which a sample is drawn.
- For example if you are interested to find out the average age of Indian then some census data can be thought of sampling frame
- If you are interested to find out average income of a city, probably city telephone directory can act as sampling frame

Importance of Sampling Frame

"In many practical situations the frame is a matter of choice to the survey planner, and sometimes a critical one. Some very worthwhile investigations are not undertaken at all because of the lack of an apparent frame; others, because of faulty frames, have ended in a disaster or in cloud of doubt."

— *Raymond James Jessen*

Types of sampling frames

- The most straightforward type of frame is a list of elements of the population (preferably the entire population) with appropriate contact information.
  - For example, in an opinion poll, possible sampling frames include an electoral register or a telephone directory.

- Not all frames explicitly list population elements; some list only 'clusters'.
  - For example, a street map can be used as a frame for a door-to-door survey

Sampling frames problems

- Omission of sample.
- Inclusion of non-element.
- Repetition of sample
- In some cases, this may lead to sampling bias.

Types of Sampling

- Probability Sampling
  - Simple random sample
  - Systematic random sample
  - Stratified random sample
  - Multistage sample
  - Multiphase sample
  - Cluster sample
- Non-Probability Sampling
  - Convenience sample
  - Purposive sample
  - Quota

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.

NON PROBABILITY SAMPLING

- Any sampling method where some elements of population have no chance of selection (these are sometimes referred to as 'out of coverage'/‘ undercovered'), 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.
- Example: We visit every household in a given street, and interview the first person to answer the door. In any household with more than one occupant, this is a nonprobability sample, because some people are more likely to answer the door (e.g. an unemployed person who spends most of their time at home is more likely to answer than an employed housemate who might be at work when the interviewer calls) and it's not practical to calculate these probabilities.

SIMPLE RANDOM SAMPLING

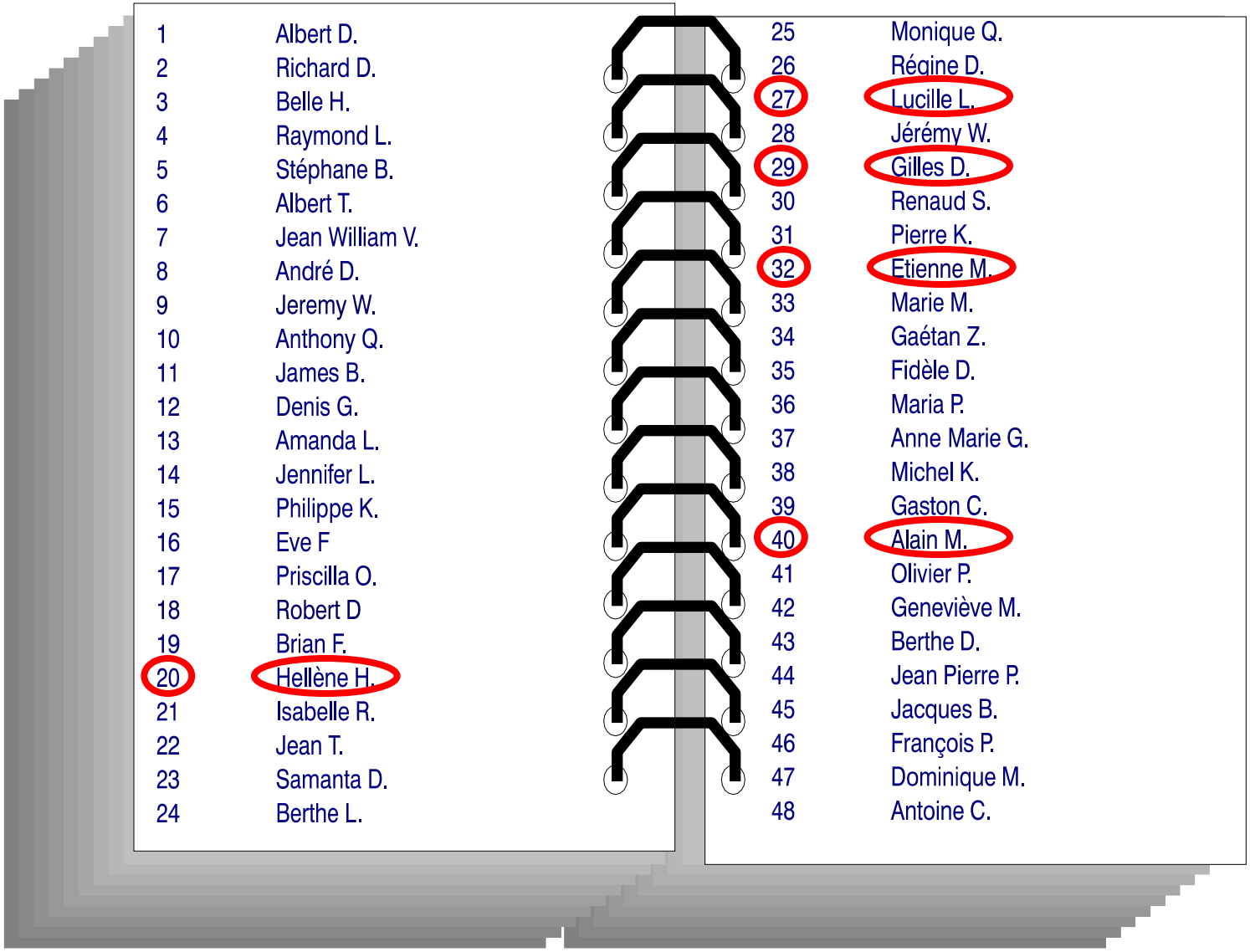
- Applicable when population is small, homogeneous & readily available
- All subsets of the frame are given an equal probability. Each element of the frame thus has an equal probability of selection.

Example: The American Community Survey is an example of simple random sampling. In order to collect detailed data on the population of the US, the Census Bureau officials randomly select 3.5 million households per year and use a variety of methods to convince them to fill out the survey.

Disadvantages

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

Simple random sampling



SYSTEMATIC SAMPLING

- **Systematic sampling** relies on arranging the target population according to some ordering scheme and then selecting elements at regular intervals through that ordered list.
- Systematic sampling involves a random start and then proceeds with the selection of every kth 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 kth element in the list.
- A simple example would be to select every 10th name from the telephone directory (an 'every 10th' sample, also referred to as 'sampling with a skip of 10').

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.

SYSTEMATIC SAMPLING……

ADVANTAGES:

- Sample easy to select
- Suitable sampling frame can be identified easily
- Sample evenly spread over entire reference population

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

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.
- 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.

STRATIFIED SAMPLING……

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

Drawbacks

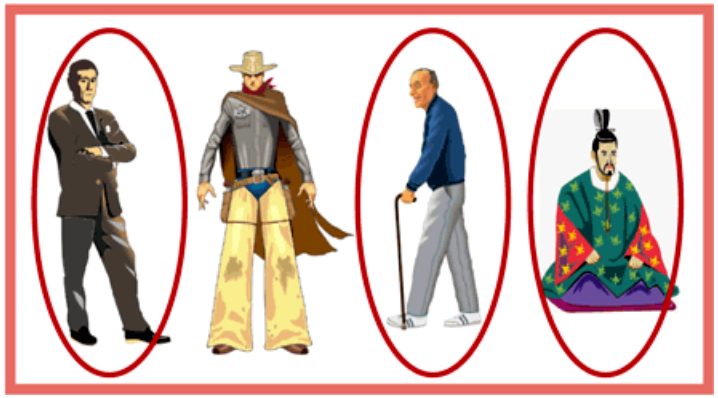
- First, sampling frame of entire population has to be prepared separately for each stratum
- Second, when examining multiple criteria, stratifying variables may be related to some, but not to others, further complicating the design, and potentially reducing the utility of the strata.
- Finally, in some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

STRATIFIED SAMPLING…….

Draw a sample from each stratum

Women

Men



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 talukas, villages.
- Then third stage units will be houses.
- All ultimate units (houses, for instance) selected at last step are surveyed.

MULTI PHASE SAMPLING

- Part of the information collected from whole sample & part from subsample.

- In Tb survey MT in all cases – Phase I
- X –Ray chest in MT +ve cases – Phase II
- Sputum examination in X – Ray +ve cases - Phase III

- Survey by such procedure is less costly, less laborious & more purposeful

CLUSTER SAMPLING

- 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.

CLUSTER SAMPLING…….

Advantages :

- Cuts down on the cost of preparing a sampling frame.
- This can reduce travel and other administrative costs.

Disadvantages:

- sampling error is higher for a simple random sample of same size.

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

Non-Probability Sampling

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

CONVENIENCE SAMPLING

- Sometimes known as grab or opportunity sampling or accidental or haphazard sampling.
- A type of nonprobability 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.
- 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

Network or Snowball Sampling

- Takes advantage of social networks to get the sample
- One person in the sample asks another to join the sample, and so on.

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