SAMPLING METHODS

LEARNING OBJECTIVES

- **Learn the reasons for sampling**
- **Develop an understanding about different sampling methods**
- **Distinguish between probability & non probability sampling**
- **Discuss the relative advantages & disadvantages of each sampling methods**

What is research?

- “Scientific research is systematic, controlled, empirical, and critical investigation of natural phenomena guided by theory and hypotheses about the presumed relations among such phenomena.”
  - Kerlinger, 1986

- *Research is an organized and systematic way of finding answers to questions*

Important Components of Empirical Research

- Problem statement, research questions, purposes, benefits
- Theory, assumptions, background literature
- Variables and hypotheses
- Operational definitions and measurement
- Research design and methodology
- Instrumentation, sampling
- Data analysis
- Conclusions, interpretations, recommendations

SAMPLING

- 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.

SAMPLING……

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

SAMPLING…….

- 3 factors that influence sample representative-ness
  - 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

Who do you want to generalize to?

The Theoretical Population

What population can you get access to?

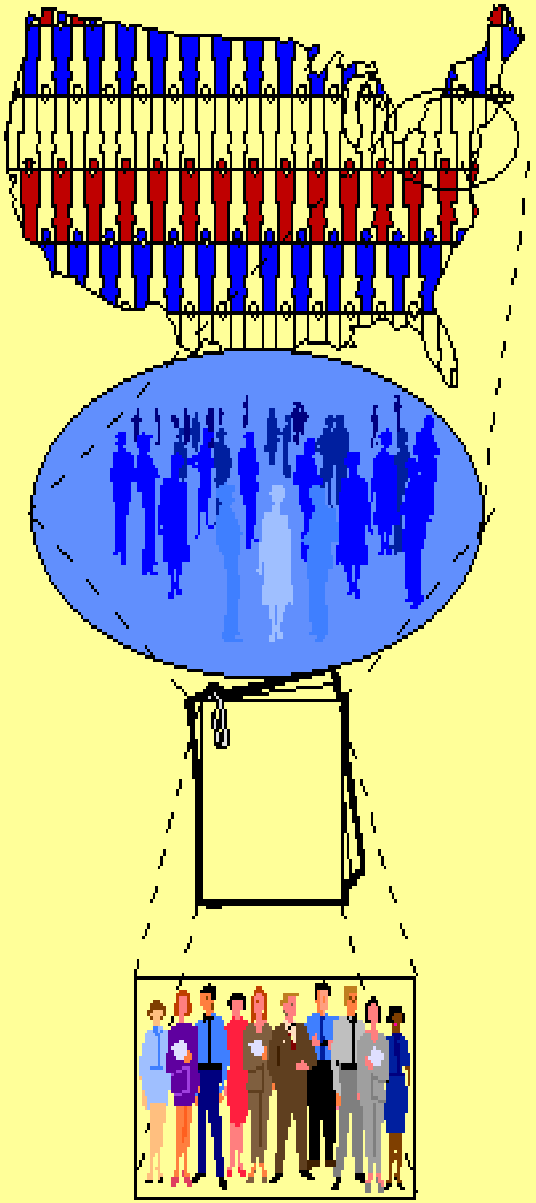
The Study Population

How can you get access to them?

The Sampling Frame

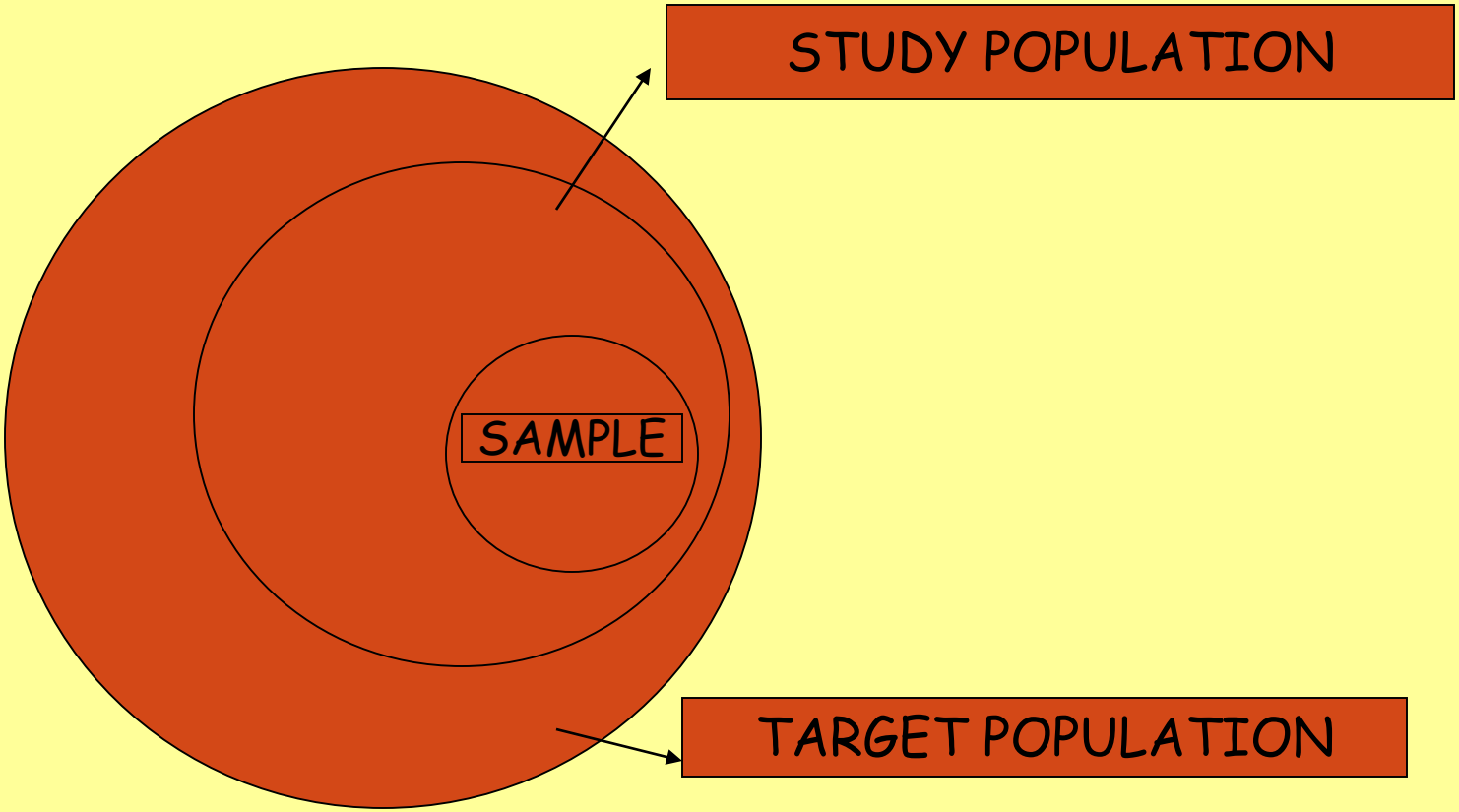
Who is in your study?

The Sample



SAMPLING BREAKDOWN

SAMPLING.....



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

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.

SAMPLING FRAME

- In the most straightforward case, such as the sentencing of a batch of material from production (acceptance sampling by lots), it is possible to identify and measure every single item in the population and to include any one of them in our sample. However, in the more general case this is not possible. There is no way to identify all rats in the set of all rats. Where voting is not compulsory, there is no way to identify which people will actually vote at a forthcoming election (in advance of the election)
- As a remedy, we seek a *sampling frame* which 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

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…….

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

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

- *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*.

NONPROBABILITY SAMPLING…….

- 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.

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.

SIMPLE RANDOM SAMPLING……..

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

- **Disadvantages**
- If sampling frame large, this method impracticable.
- 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.

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 *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.
- 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** 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

STRATIFIED SAMPLING…….

Draw a sample from each stratum

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.
- Often used to evaluate vaccination coverage in EPI

CLUSTER SAMPLING…….

- **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 = 4th cluster.
  - Last or 30th cluster = 29th cluster + sampling interval

CLUSTER SAMPLING…….

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.

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.

MULTISTAGE SAMPLING……..

- 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.
- 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.

QUOTA SAMPLING

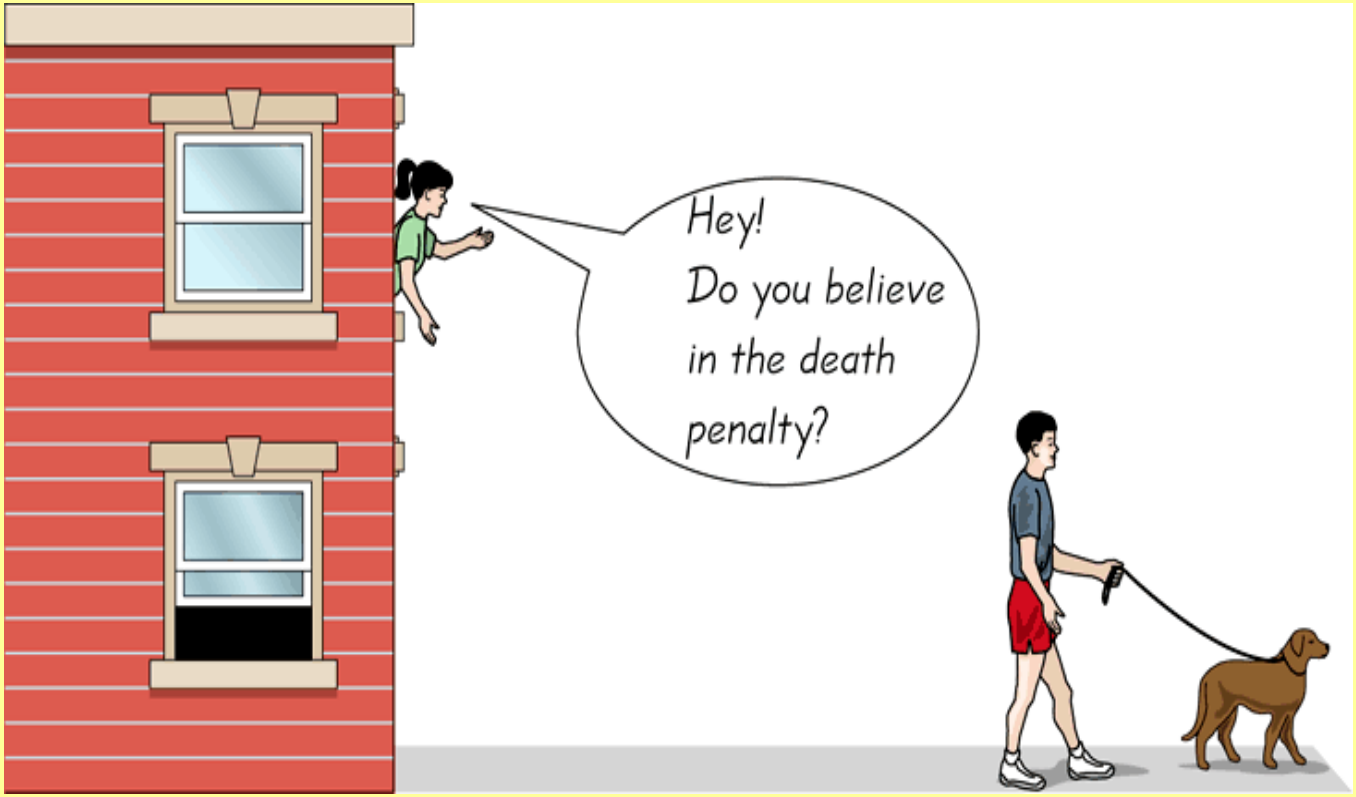
- The population is first segmented into mutually exclusive sub-groups, just as in stratified sampling.
- Then judgment used to select subjects or units from each segment based on a specified proportion.
- For example, an interviewer may be told to sample 200 females and 300 males between the age of 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, where existing study subjects are used to recruit more subjects into the sample.

CONVENIENCE SAMPLING…….

- Use results that are easy to get



35

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

Contd.

- Example: Suppose we have six schools with populations of 150, 180, 200, 220, 260, and 490 students respectively (total 1500 students), and we want to use student population as the basis for a PPS sample of size three. To do this, we could allocate the first school numbers 1 to 150, the second school 151 to 330 (= 150 + 180), the third school 331 to 530, and so on to the last school (1011 to 1500). We then generate a random start between 1 and 500 (equal to 1500/3) and count through the school populations by multiples of 500. If our random start was 137, we would select the schools which have been allocated numbers 137, 637, and 1137, i.e. the first, fourth, and sixth schools.
- The PPS approach can improve accuracy for a given sample size by concentrating sample on large elements that have the greatest impact on population estimates. PPS sampling is commonly used for surveys of businesses, where element size varies greatly and auxiliary information is often available - for instance, a survey attempting to measure the number of guest-nights spent in hotels might use each hotel's number of rooms as an auxiliary variable. In some cases, an older measurement of the variable of interest can be used as an auxiliary variable when attempting to produce more current estimates.