

Survey Methods 101

Professor Khandker Nurul Habib, PhD, PEng
Department of Civil Engineering
University of Toronto

October 2nd Workshop on Household Travel Survey
Data Management Group, UTTRI
University of Toronto

Outlines

- Introduction
- Terminology
- Basic stages of any survey
- Subject representation, temporal perspective and information perspective
- Mode of survey
- Quality, validity and errors
- Construct and measurement
- Survey and data; inference and error
- Sample and weighting
- Ethical consideration
- Concluding remarks

Introduction: Travel Demand Data

Who

- Residents , Non-Residents, visitors

Why

- Engaging in activities: Work/school, shopping, social, recreational, other

Where

- Trip origin and destination locations

How

- Mode of transportation: single mode, multi-mode; Routes

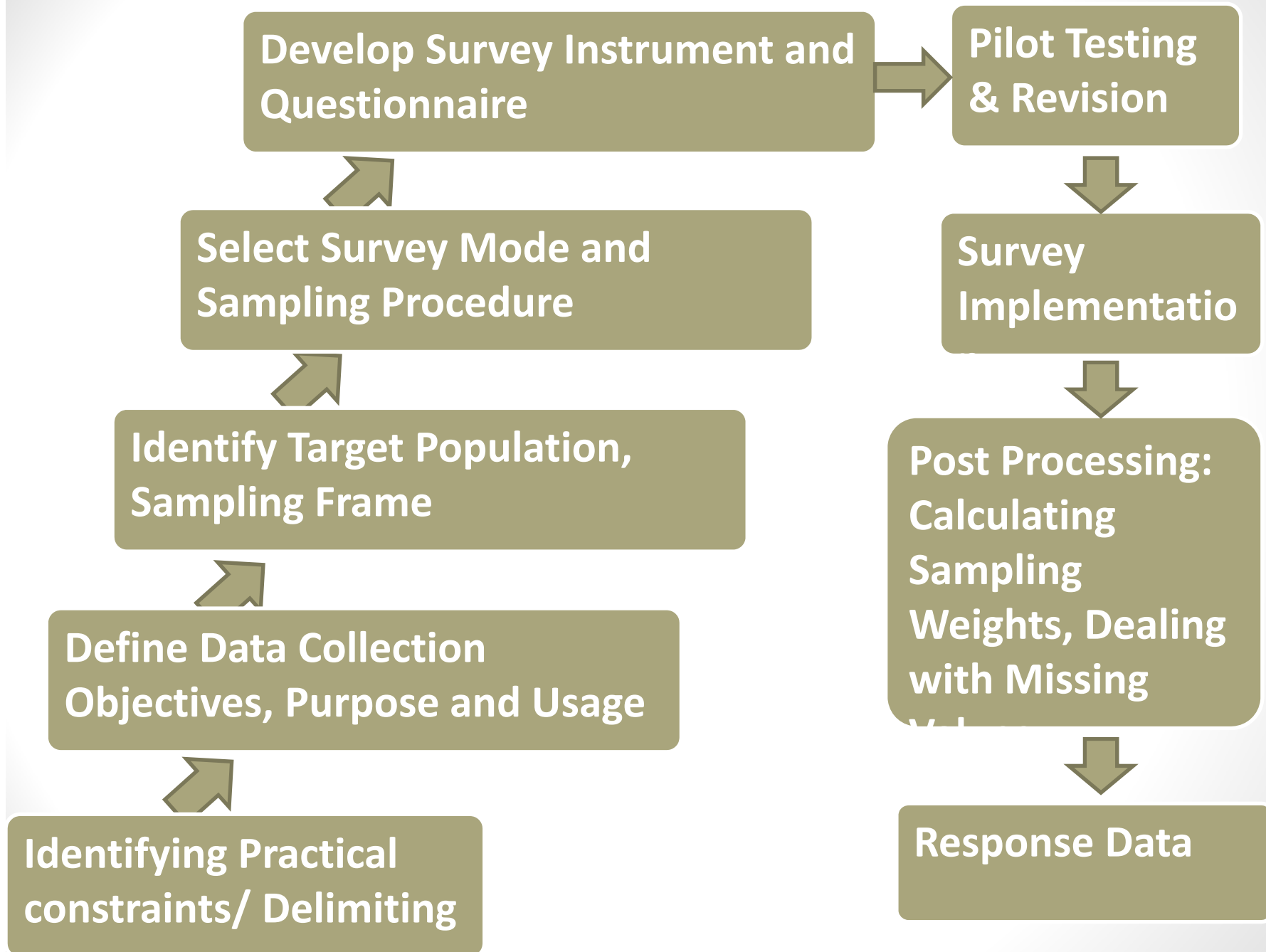
When

- Time of day, day of week, month, year, etc.

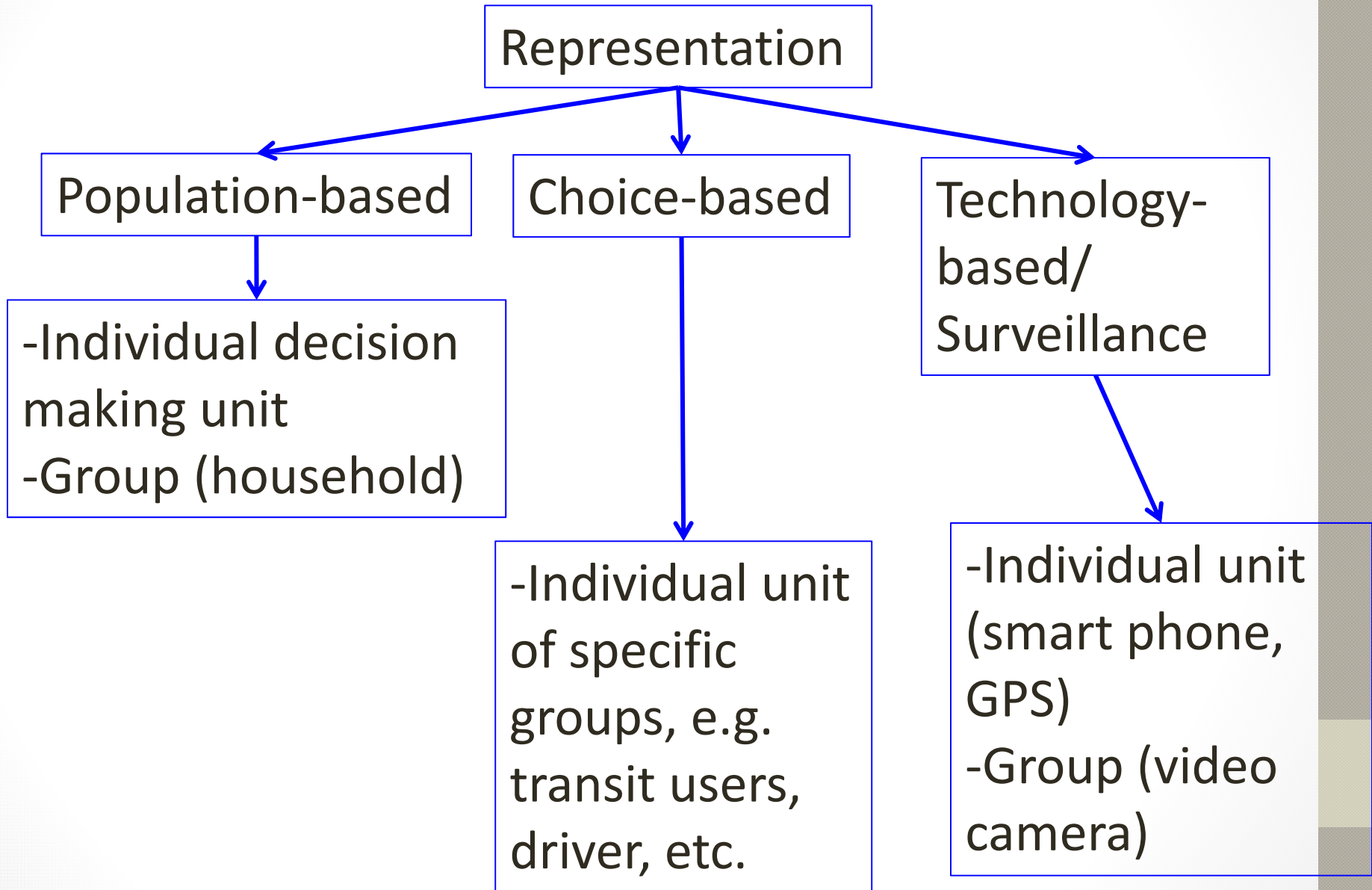
Terminology

- **Population:** target population of the study area
- **Sample frame:** complete list of contact information of the population of the study area. Such as telephone directory
- **Sample:** a subset of population from the sample frame who are contacted for the survey
- **Respondents:** individuals of the sample who responds
- **Instrument:** A single question/observation that is used to measure a specific information
- **Questionnaire:** A set of instruments that makes the whole survey
- **Survey mode:** The method/means used to present the questionnaire to the respondents, e.g. telephone, web, face-to-face, etc.
- **Construct:** Information that are to be collected
- **Measurement:** Quantifying the information

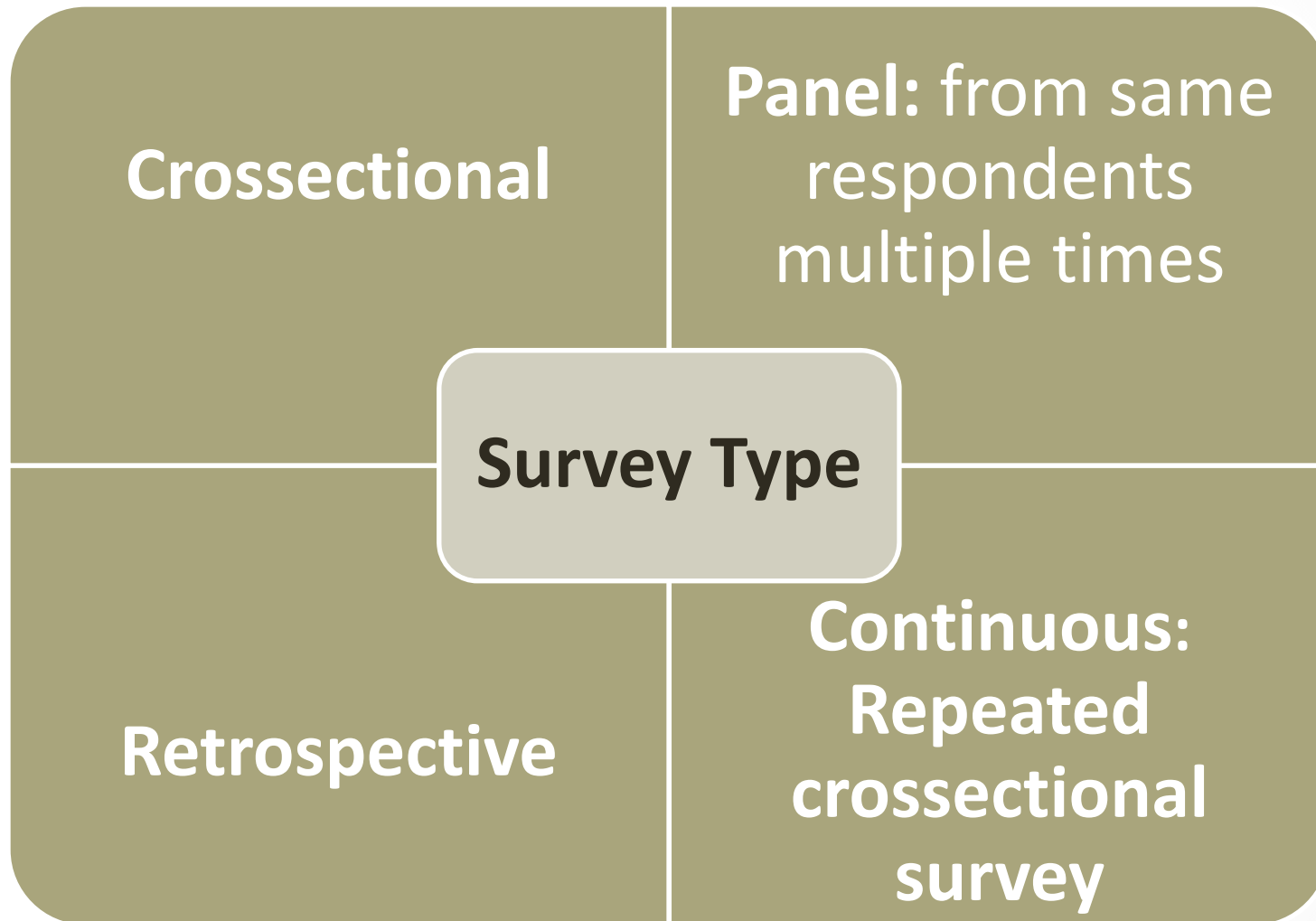
Basic Stages of any Survey



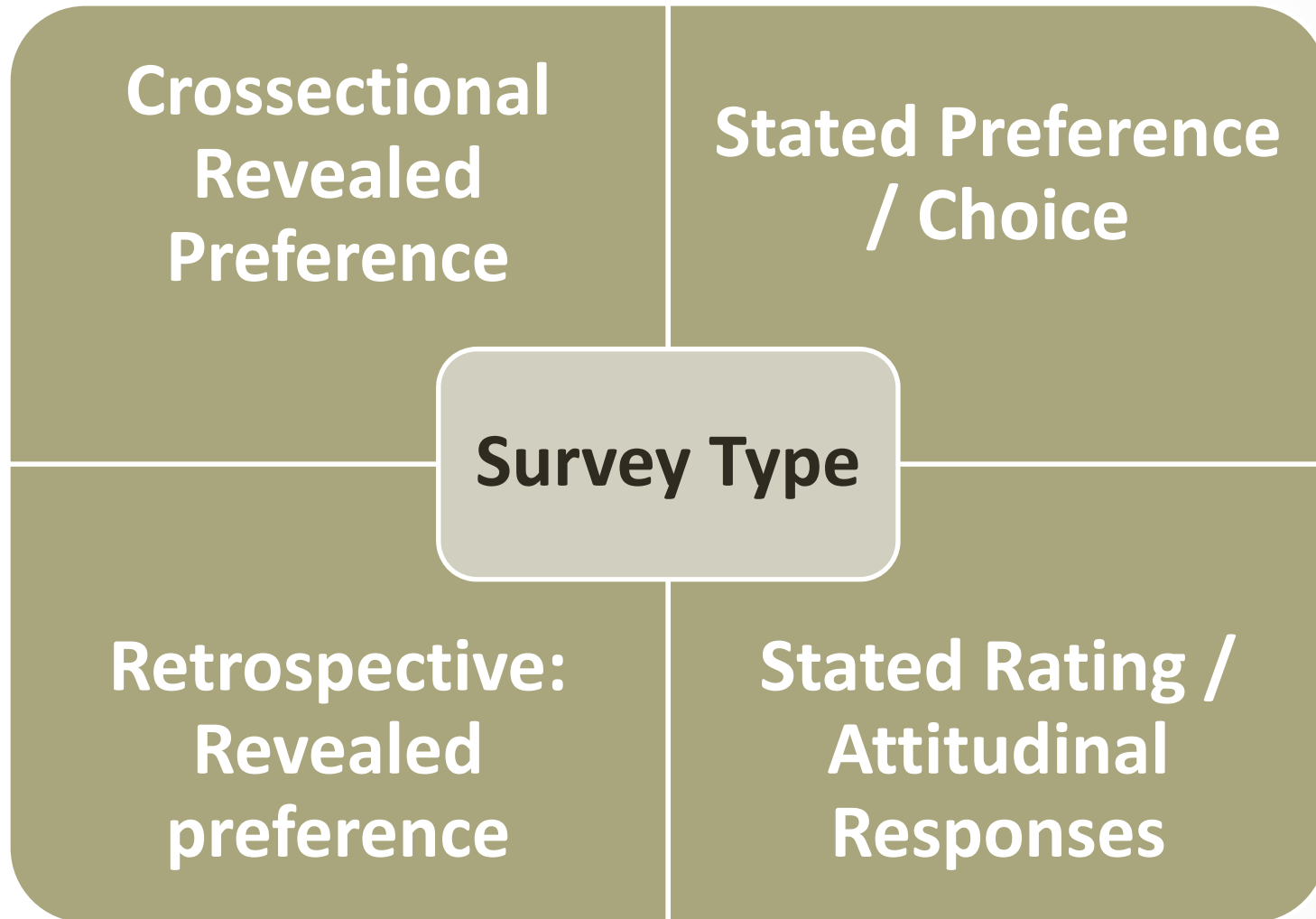
Subject Representation



Temporal Perspective



Information Perspective



Mode of Survey

1. Face-to-face
2. Mail-based
3. Telephone
4. Web-based tools, Social media, etc.
5. Surveillance: cell phone, GPS, remote sensing technology
6. Smart card, transit pass, debit/credit card usage, etc
7. Mixed modes

Quality-Validity-Error of Sample Survey

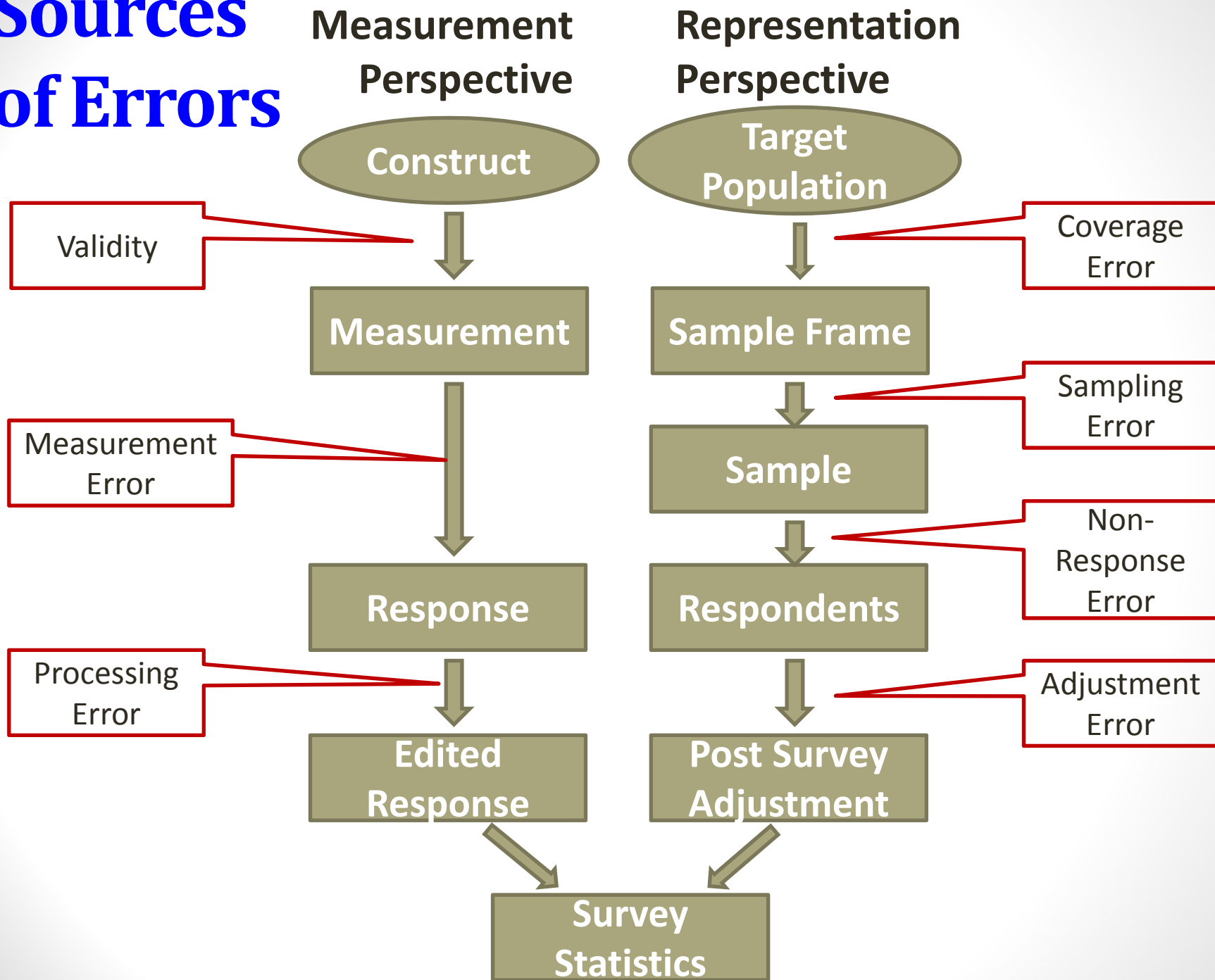
- **Quality** refers to the representation: How well sample represents the population through sample statistics
- **Validity** refers to the validity of measurements
- **Errors:** Bias (Systematic) & Variance
- Survey quality and validity are influenced by
 - ✓ subject representation
 - ✓ Temporal perspective
 - ✓ Information perspective
 - ✓ Mode of data collection

Construct & Measurement

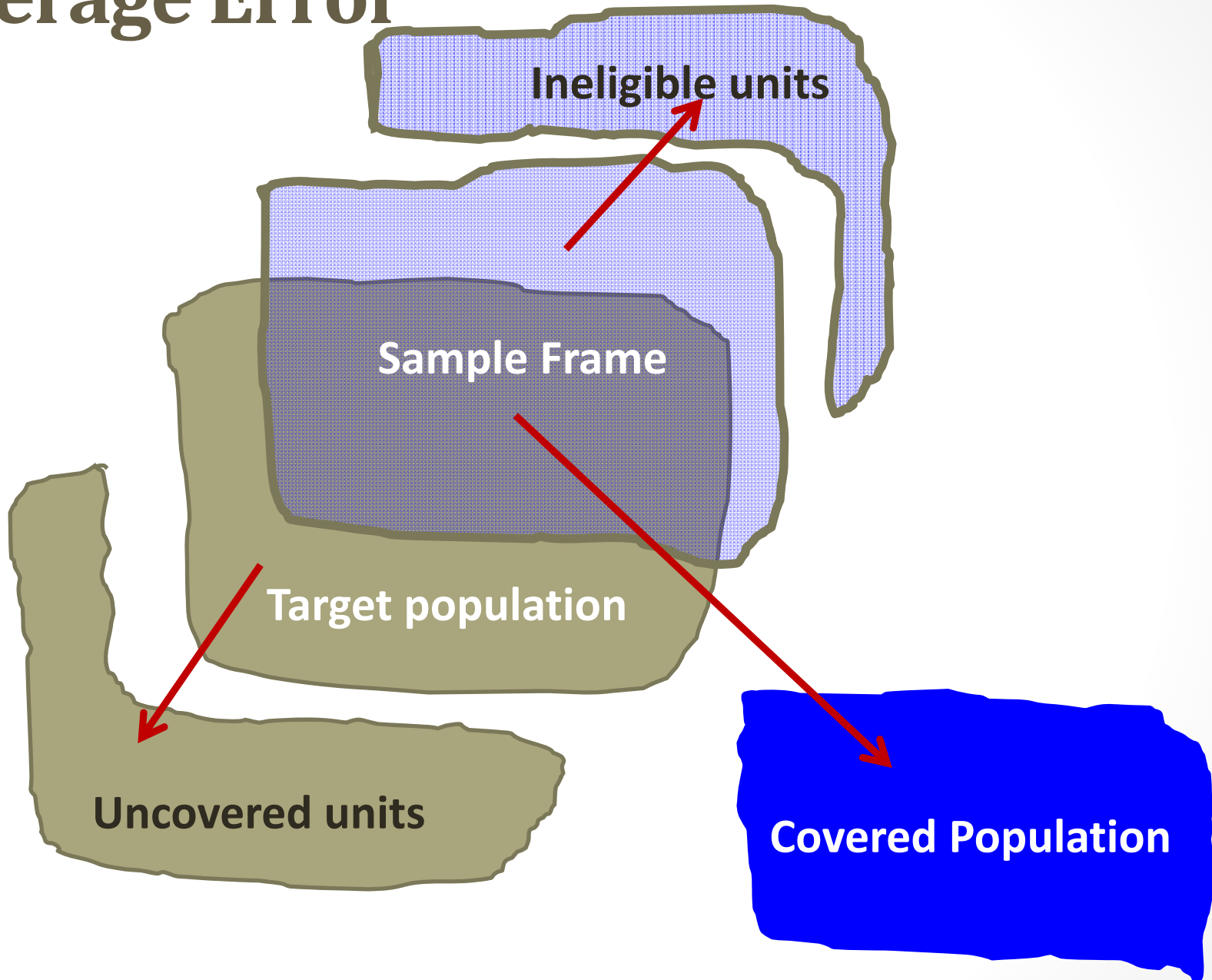
- **Construct** is the information that we need to know:
Travel demand
 - ☐ **Measurement :**
 - ✓ number of trips made
 - ✓ trip purposes/activity types
 - ✓ trip origin-destination
 - ✓ travel mode
 - ✓ travel route
 - ✓ time of day
 - ✓ with-whom travelling

- Constructs are latent and subjective
 - ✓ Refers different thing to different people
 - ✓ Single measurement may not adequately capture the construct

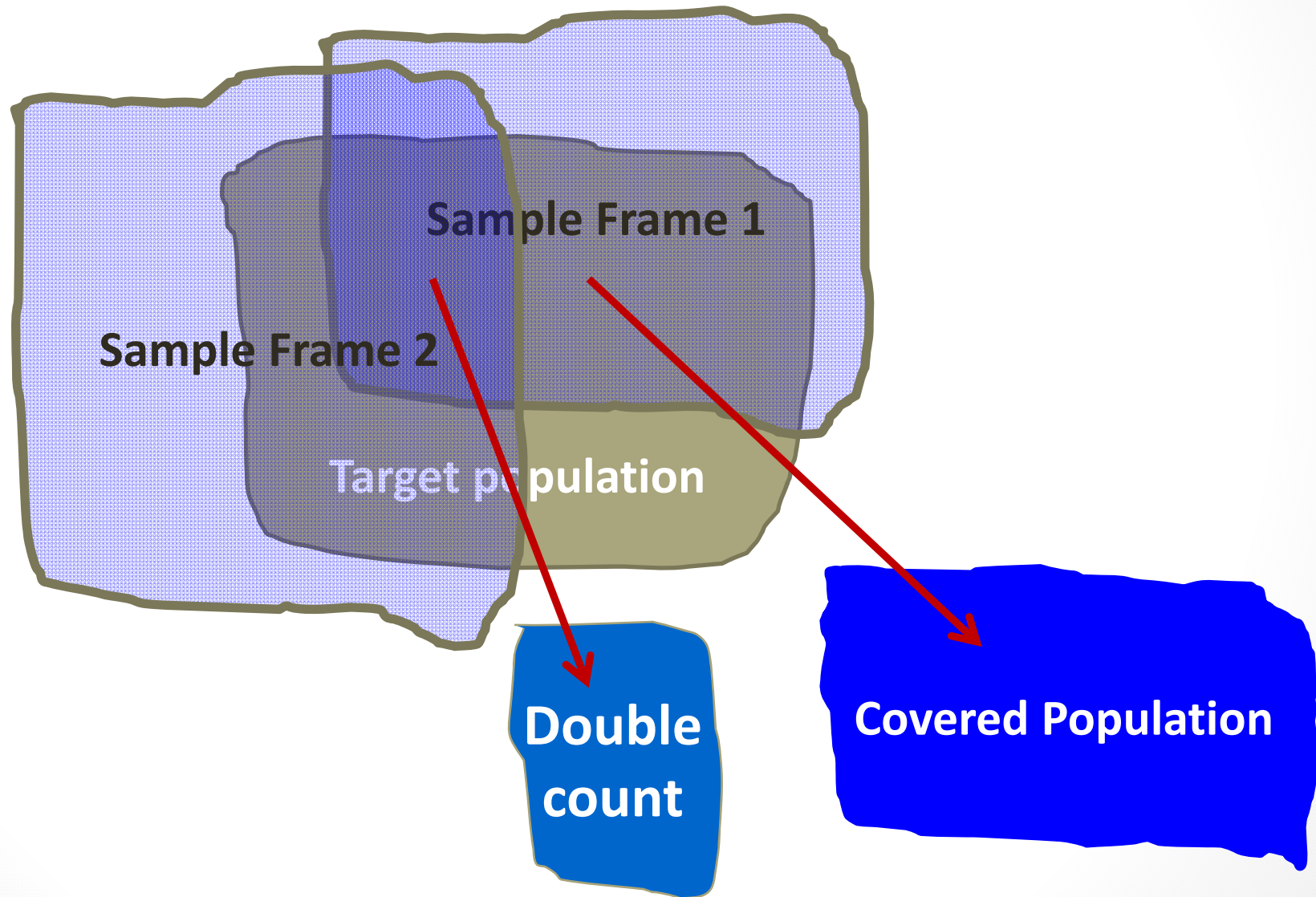
Sources of Errors



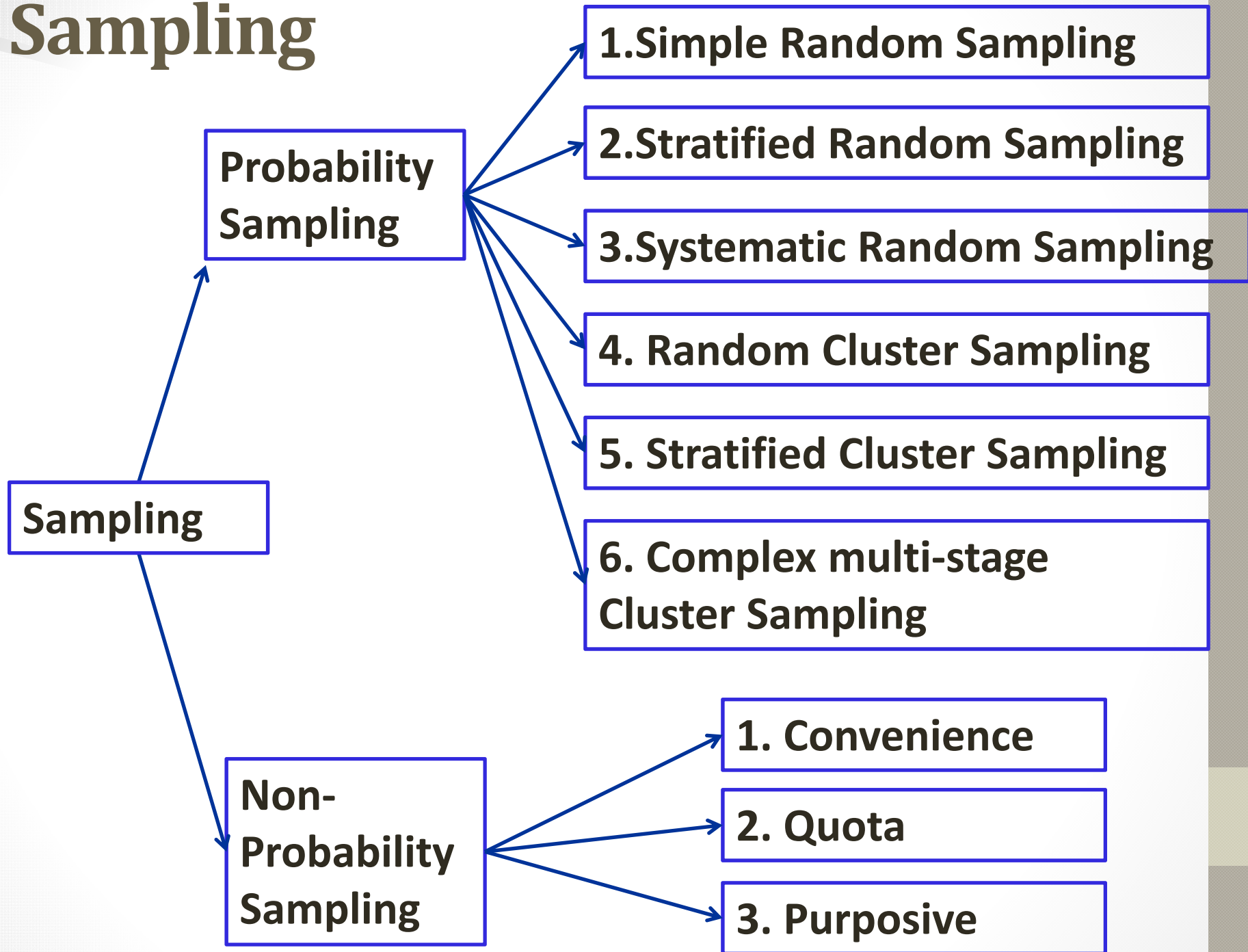
Coverage Error



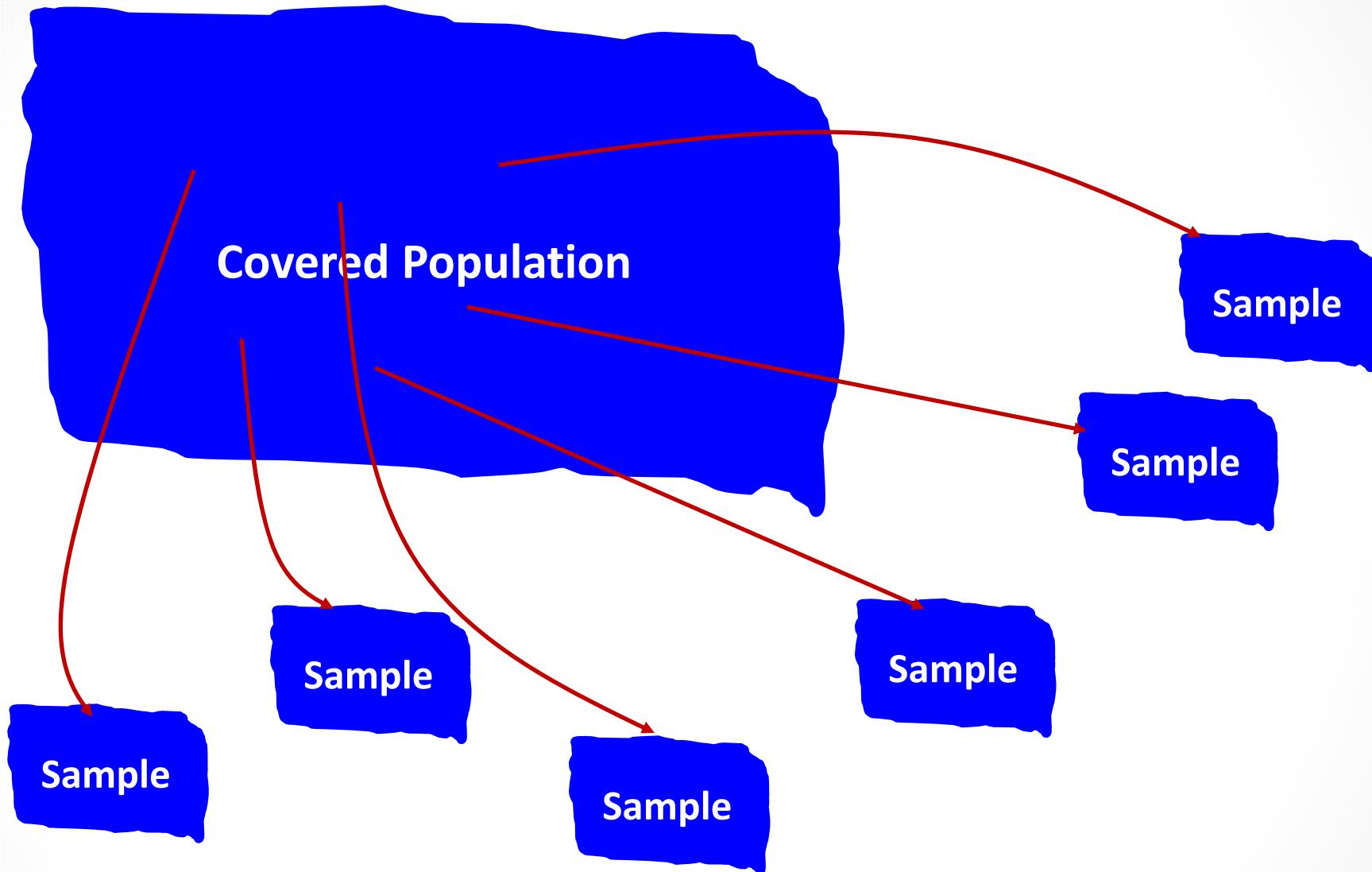
Coverage Error



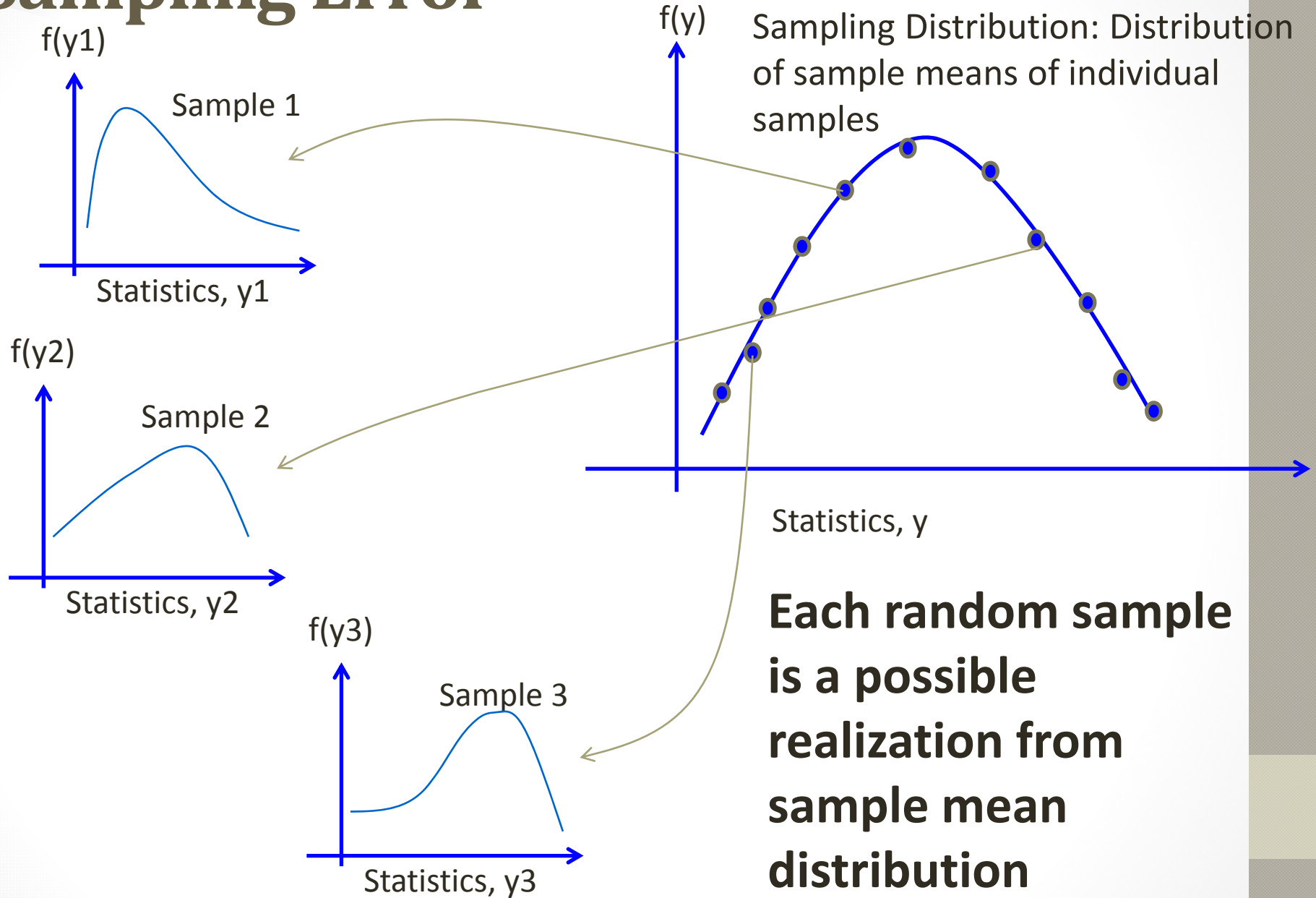
Sampling



Sampling from Sample Frame(s)



Sampling Error



Sampling Error

- Sampling error is deliberately created and large sample size can eliminate many errors
- Non-response rate adds up to sampling error
- Sample size is determined based on:
 1. The variability (across the population) in the parameters to be measured.
 2. The degree of precision required for the parameter estimated.
 3. The population (sub-population) size.

Sampling Error

- Two types of **sampling error**:
 - ✓ Sampling bias: arises when some members of the sampling frame is not given a chance to be asked
 - ✓ Sampling variance : it rests on the notion of variation of responses across those who surveyed
- If the average sample mean is equal to the mean of the sampling frame, then there is no sampling bias for the mean.
- Sample variance is 'zero' only in population with constant value of variable

Sampling Error

The **Extent of Error** due to **sampling** is a function of four basic principles:

1. Whether all sampling frame elements have known, non-zero chance of selection into the sample.
 2. Whether the sample is designed to control the representation of key subpopulation in the sample.
 3. Whether individual elements are drawn directly and independently or in groups.
 4. How large a sample of elements is selected.
- Non-response errors arises when the values of statistics computed based only on respondent data different from those based on the entire sample frame.

Weighting

- Attaching a weight to each observation to expand to represent the whole target population
- Helps compensate for unit non-response
 - ✓ weights are adjusted to align survey results with known population figures
- Survey weights should be calculated for responding unit and is needed to adjust for varying probabilities of selection
- Varying selection probabilities arise from
 - ✓ Stratification
 - ✓ Selecting one person per household
 - ✓ Double sampling, e.g. for booster samples
- Differential weights induce adjustment error

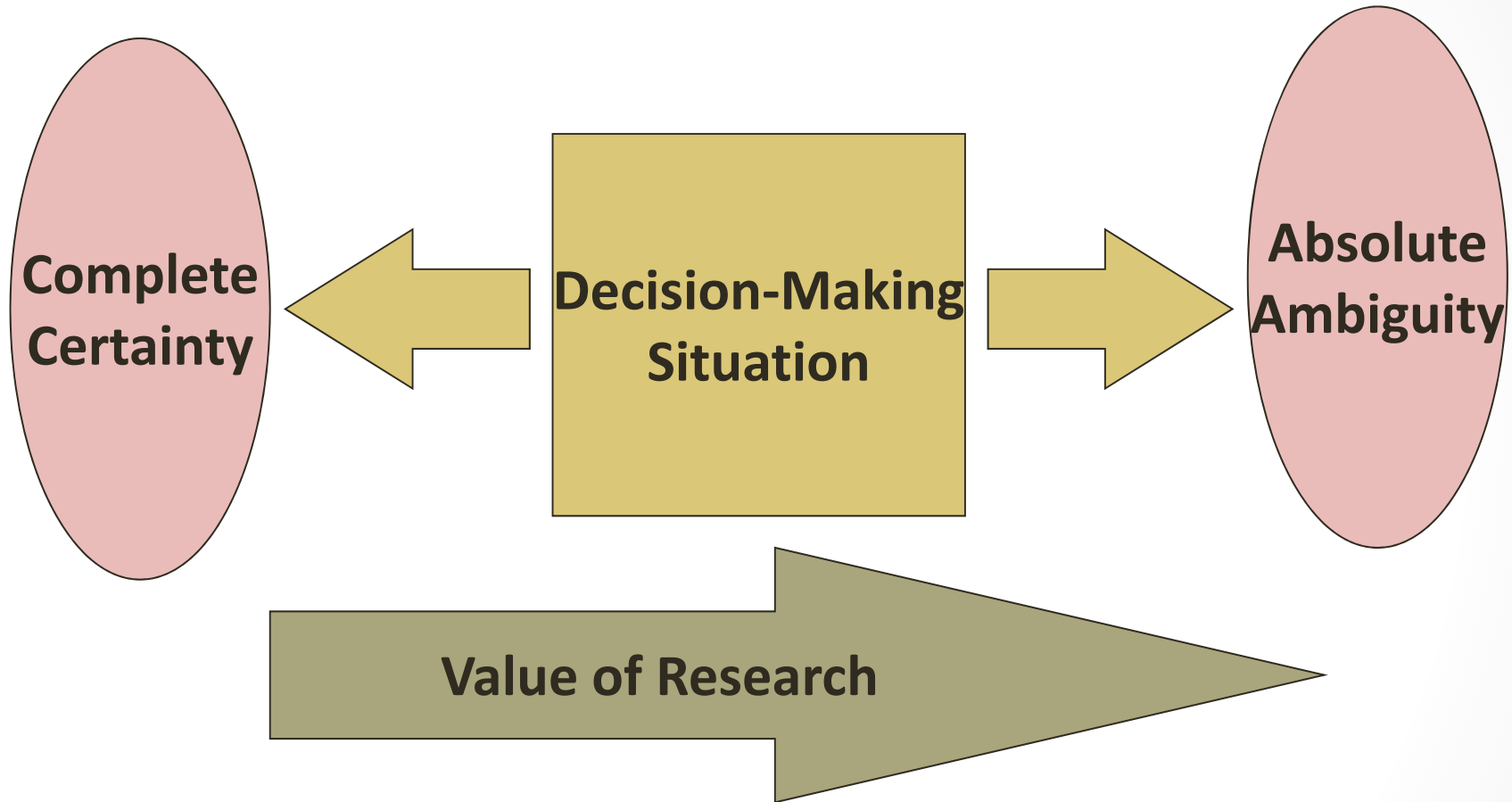
Precision and Accuracy

- Precise survey → displays **repeatability** → yielding the same response if administered repeatedly under same conditions. :
 - ✓ Can be increased by increasing sample size by reducing the possibility of unobserved members .
- Accurate survey → survey is aimed at correct sample of the correct population.:
 - ✓ Can be increasing by ensuring that members are not systematically eliminated in the frame and the sample is truly random.
- We need equal attention to accuracy and precision → tendency to overcome sampling bias by increasing sample size only may not work all the time.
- It is important to identify a sample frame correlated with survey objective:
 - ✓ **Correlated**: Motor vehicle registration list as the frame of car ownership.
 - ✓ **Not Directly Correlated**: Telephone directory as Sample frame for travel survey.

Studying Human and Society

- The object of study is not "dead" → reflects and responds to new knowledge
- The object of study is not static but in constant flux → temporality of knowledge
- Researcher is part of the object of study (society) → objectivity problematic.
- Measurement problematic?

Data on Human Decisions



Ethical Considerations

- No harm done
- Informed consent
- Anonymity
- Confidence

Ethical Considerations

Subject/Interviewee/Respondents

- Truthfulness in giving information to the researcher if a research subject or respondent gives his or her consent to participate in a research study
- Sustained cooperativeness with the researcher throughout the course of the research study
- Adhere to responsibility if informed consent is given to the researcher
- State any constraints or limitations in advance

Ethical Considerations

Surveyor/Data users/Researchers

- No deception, be forthright and do not conceal the true purpose of the research
- Maintain objectivity, courtesy and high professional standards through scientific process
- No falsification, alteration or misrepresentation of data for political or other purposes
- Protect the confidentiality of the research subjects and research sponsors
- No faulty conclusions
- No inclusion or use of information or ideas contained in competing research proposals

Ethical Considerations

Sponsors/Funding Agencies:

- Avoid manipulation and influencing with a view to discrediting individuals or organizations
- The conclusions drawn from research work should be consistent with the data and not influenced by other undesirable conditions or motives
- Observe the confidentiality of the research subjects and researcher
- Avoid Advocacy objectives!

Concluding Remarks

- Research is necessary to disentangle effects of survey mode on data validity and quality
- Challenges of collecting temporal and spatial details without necessarily overburdening the respondents needs to be sorted out
- Research is necessary to reconstructing household level effects from individual-based survey
- Research is necessary for efficient combinations of revealed, retrospective, stated and attitudinal question