AGRODEP

Household survey data course

Dakar, 8-10 October 2012

Sample design

IFPRI®

HarvestChoice

BETTER CHOICES, BETTER LIVES

THE WORLD BANK

LSMS

Living Standards Measurement Study

Outline

- Purpose, concepts, and definitions
- Simple Random Sampling

  -Sampling and Non-sampling errors, confidence intervals, sample size, population size, sampling for rare events
- Multi-stage and Cluster Sampling
- Stratification
- Weighting
- Design Effects
- Sampling for Impact Evaluations
- Group Exercise

What we are **not** going to talk about…

- Being an expert in sample design
- Non-random designs (matching, RDD, etc)

Purpose of sampling

- Why sampling?
  - Saves **cost** compared to full enumeration
  - Easier to control **quality** of sample
  - More **timely** results from sample data
  - Measurement can be destructive

Sampling Concepts and Definitions

Unit of analysis

- The level at which a measurement is taken

- Most common units of analysis are persons, households, farms, and economic establishments

Sampling Concepts and Definitions

Target population or universe

- The complete collection of *all the units of analysis* to study.

- Examples: population living in households in a country; students in primary schools

Sampling Concepts and Definitions

Sampling frame

- List of all the units of analysis whose characteristics are to be measured
- Comprehensive, non-overlapping and must not contain irrelevant elements
- Should be updated to ensure complete coverage
- Examples: list of establishments; census; civil registration

Sampling Concepts and Definitions

Parameter / Estimate

- Objective of sampling is to estimate parameters of a population
- Quantity computed from all N values in a population set
- Typically, a descriptive "true value" measure of a population, such as mean, variance
  - Poverty rate, average income, etc.

Sampling Concepts and Definitions

Unbiased Estimator

- **Estimator** - mathematical formula or function using sample results to produce an estimate for the entire population

$$\hat{\theta} = \hat{\theta}(X_1, X_2, ..., X_n)$$

- When the mean of individual sample estimates equals the population parameter, then the estimator is unbiased
- Formally, an estimator is unbiased if the expected value of the (sample) estimates is equal to the (population) parameter being estimated (where k is the number of experiments).

$$\frac{\hat{\theta}_1 + \hat{\theta}_2 + ... + \hat{\theta}_k}{k} \xrightarrow[k \to \infty]{} \theta$$

Sampling Concepts and Definitions

Central Limit Theorem

- In making inference from samples we rely on CLT

  "For reasonably large sample sizes (30 and above), the distribution of possible sample means will tend towards a normal distribution"

- Properties of sample estimators derived theoretically from pattern of results from repeated samples n times (sampling distribution)

Types of Sampling

- Random
- Non-random
  - Purposive
  - Quota
  - Accidental
  - Networking

Non-Random Sampling

- Purposive or judgmental sampling – units selected based on special criteria
  - select areas/units for conducting survey pilots
  - " wise old men"
  - not possible to make inferences to population
- Quota sampling
  - Used for convenience, sometimes incorrectly treated as a random sample
  - Select % within different population groups; selection left to enumerators
  - Sample does not represent population, estimates will be biased

Non-Random Sampling (cont'd)

- Accidental
  - Pick as you go
- Networking
  - Hard to find groups
  - People connected (migrants abroad)

- …can be due to past sampling! -> Panel Survey

Panel Surveys can measure change better

It seems that $Y_{2001} > Y_{2005}$ but…

…both measures are affected by sampling errors (e_{2001} et e_{2005})

The error of the difference $Y_{2005} - Y_{2001}$ is…

…$\sqrt{(e^2_{2001} + e^2_{2005})}$ if the two samples are independent

…only $\sqrt{(e^2_{2001}+e^2_{2005}-2\varrho[Y_{2001},Y_{2005}])}$ if the sample is the same

Advantages and disadvantages of panels

- **Analytical advantages**
  - Can measure changes better
  - Permit understanding better why things changed
  - Permits correlating past and present behavior

- **Analytical disadvantages**
  - Become progressively less representative of the population

- **Practical advantages**
  - No sampling design needed for the second and subsequent surveys

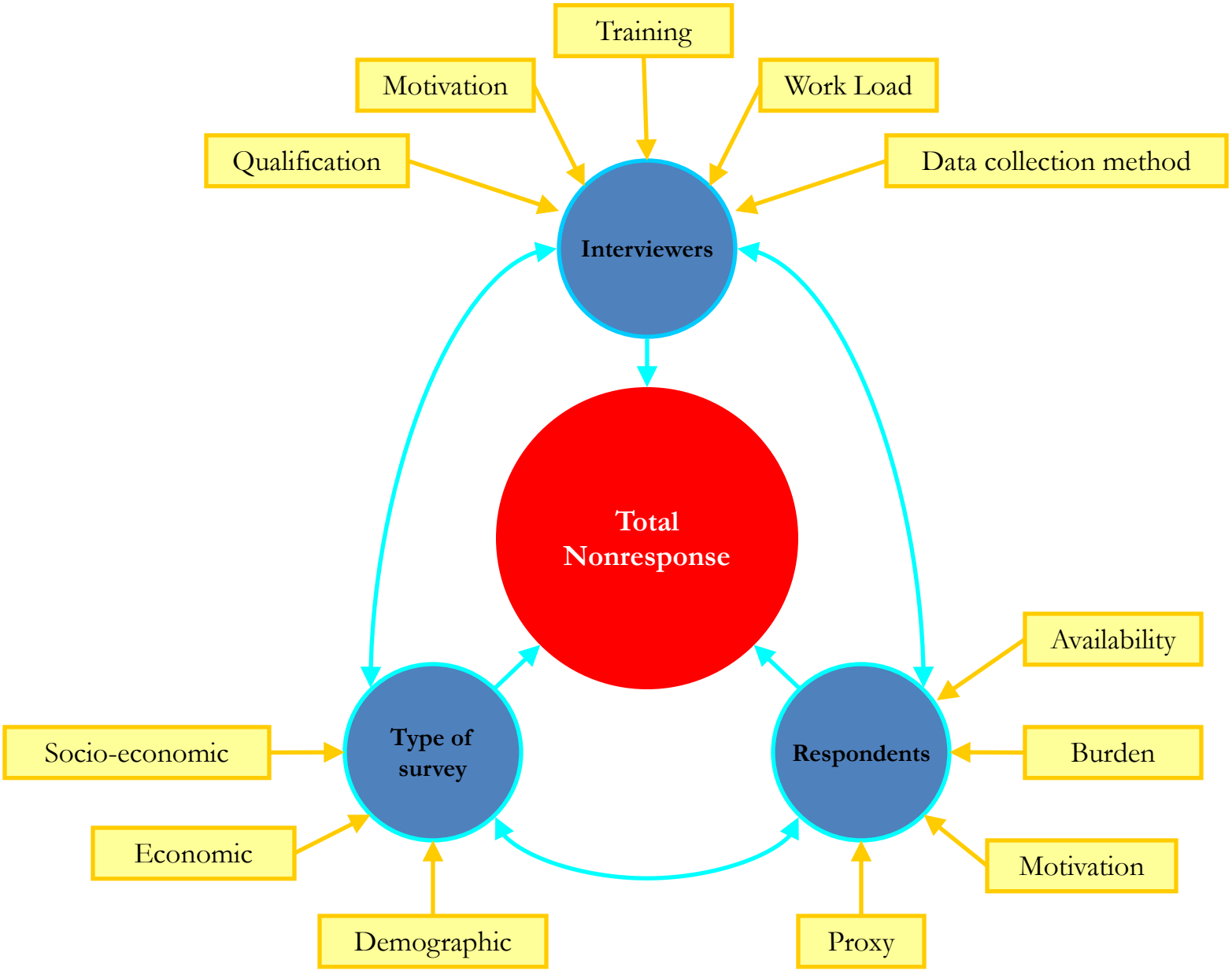
- **Practical disadvantages**
  - Sample attrition
  - Much harder to manage
  - Better to design them prospectively rather than in afterthought

Nonresponse

- Possible solutions…
  - ❑ Replace non-respondents with similar households
  - ❑ Increase the sample size to compensate for it
  - ❑ Use correction formulas
  - ❑ Use imputation techniques (hot-deck, cold-deck, warm-deck, etc.) to simulate the answers of non-respondents
  - ☑ None of the above

The best way to deal with nonresponse is to prevent it

Lohr, Sharon L. *Sampling: Design & Analysis* (1999)



Source: "Some factors affecting Non-Response." by R. Platek. 1977. *Survey Methodology*. *3*. 191-214