Sampling Techniques for Development Evaluation

National Evaluation Capacities (NEC) Conference 2019
No One Left Behind

Hurghada, Egypt

October 20-24, 2019.

Bassirou Chitou, Ph. D.

Presentation Outline

- Workshop Objectives
- Module I: Basic Sampling Concepts
- Module II: Types of Random Sampling
- Module III: Multi-Stage Sampling

PART I: BASIC SAMPLING CONCEPTS

GENERAL OBJECTIVES OF THIS TRAINING

- Point out potential roadblocks or challenges to a successful evaluation;
- provide you with tools and ideas to deal with them;
- open your curiosity and wet your appetite to learn more.

SPECIFIC OBJECTIVES OF THIS TRAINING

- Correct common misconceptions;
- Understanding various types of random;
- Skills in drawing several types of samples;
- Skills in drawing multi-stage household samples;
- Skills in using several web-based and MS Excel tools.

MYTH 1 AND MYTH 2

Myth 1

- Sampling NOT important to the validity of an evaluation if data collection is very well done . **WRONG !**

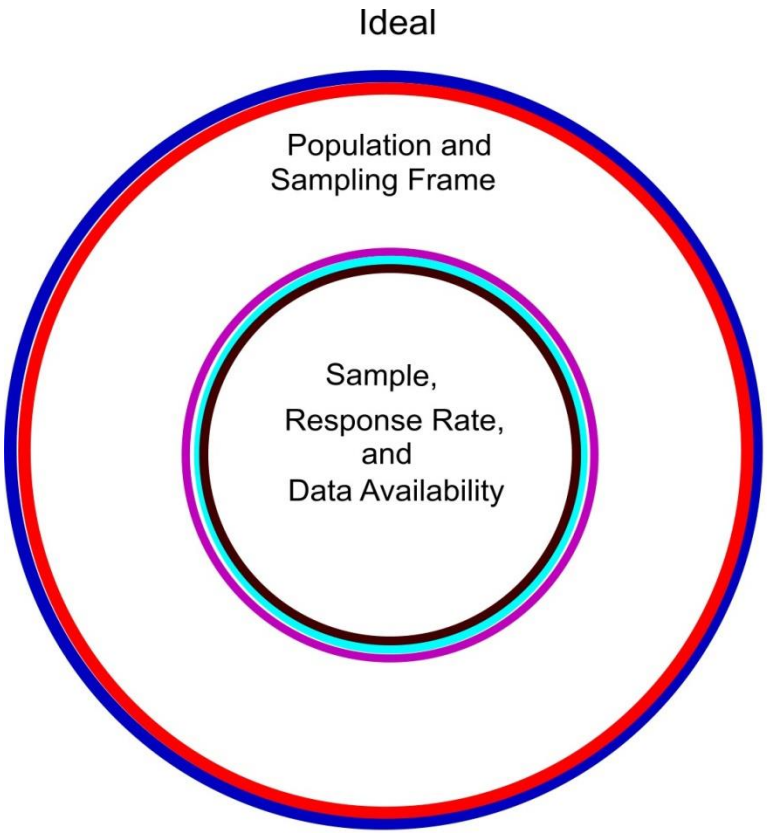
Myth 2

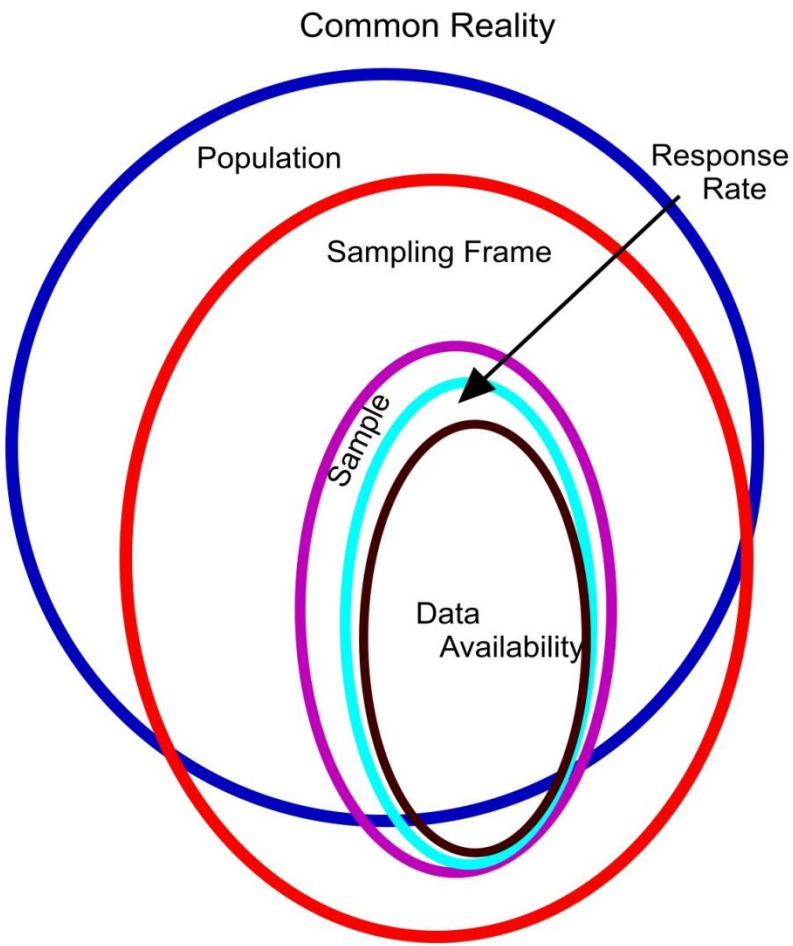
- Studies based on a sample are inherently inferior to those based on a census. **WRONG !**

Key Terms in Sampling

- Unit of Analysis
- Population (Universe)
- Sampling Frame (Frame)
- Census
- Sample
- Sampling Error
- Response Rate
- Data Availability

Relationship among Key Sampling Terms





Myth 3:

- **Non-response can be solved by over-sampling or by follow-up sampling, to achieve desired sample size.**

WRONG!

PART II:
TYPE OF RANDOM SAMPLING
PG. XX – PG. XX

Types of Random Samples

- Simple random sample
- Equal probability systematic sample (SYS)
- Random interval sample
- Stratified random sample
- Random cluster sample
- Multistage random sample

Myth 4:

- Random samples are always representative of the population from which they are drawn.

WRONG!

Myth 5:

- **Random sampling can rarely be used in development evaluations because resources usually allow only small samples and we need to assure that the sample includes units from small sectors, regions, and/or minority groups.**

WRONG!

Appendix A: World Bank Funded Operations

Show Appendix A

Myth 6:

- In development evaluation work, usually our only option is to collect data from the units convenient to the evaluator.

WRONG!

Myth 7:

- If I draw a purposeful sample to be representative of the population, I can use statistical significance and confidence intervals.

WRONG!

Table 1: Random Sample Selection Guide

Your Objective	Generally Best Type of Random Sample
Generalize to the population	
Can afford a large sample	Simple random sample
Can only afford a small sample	Stratified random sample—proportional
Expensive to access scattered units	Multi-stage random sample, starting with a random cluster sample
Sampling frame hard to enumerate	Systematic Sampling
Compare groups of unequal size in the population.	Stratified random sample-disproportional (draw the same number from each group)

Myth 8:

- **You should always draw the largest sample that is feasible.**

WRONG!

Myth 9

- If the evaluation is an important one, people will cooperate and provide the needed data.

WRONG!

SIX STRATEGIES FOR MAXIMIZING RESPONSE RATES

- Requested information will contribute to something that the respondents value—for themselves, their families, or their communities;
- Provide credible promises of confidentiality for the responses;
- Use people of influence to help solicit participation;
- Minimize the inconvenience and burden on respondents;
- Offer incentives;
- Be persistent.

PART III:
MULTI-STAGE SAMPLING

Two Reasons for Multi-Stage Sampling

- Units scattered over a large geographical area, making it time-consuming and expensive to take a simple random or stratified random sample of the units;

- there is no up-to-date sampling frame for the units of analysis but there is an up-to-date sampling frame for naturally occurring clusters of these units.

Some popular clusters

Purpose	1st stage sampling unit: Cluster	2nd stage sampling unit:	3rd stage sampling unit
Condition of bridges rated at 5 tons or greater load	Province or state	Bridges rated at 5 tons or greater load	N/A
Condom use among Female Sex Workers	Brothel, massage parlor, bar, city block	Female Sex Workers	N/A
Mechanization by small manufacturing facilities	City, town	Manufacturing facilities with less than ten employees	N/A
Youth Knowledge of HIV and Sexual Behaviors	School, household, locations where "street children" gather	Youth	N/A
Student's aspirations for secondary education	Primary schools	Fifth grade classrooms	Students
Cell phone use by street vendors	Cities	Marketplaces	Vendors

How to conduct a two-stage cluster sampling

- ***Stage 1: Selection of clusters***
  - Clusters selected using PPS
  - Listing and mapping of all SSU in each selected

- ***Stage 2: Selection of secondary sampling units* using SYS**

Advantages of two-stage cluster samples

- Two stages sampling usually offers considerable efficiency with only moderate increases in sampling errors;
- The use of residential households as the second-stage sampling unit guarantees the best coverage of the target population;
- A household listing procedure, after the selection of the first stage and before the actual data collection, provides an up-to-date sampling frame for household selection in the second stage;
- It guarantees a representative sample of the target population when there is a list of all clusters to be sampled but not a list of all target individuals within each cluster

Procedures for PPS sampling (1)

- **Step 1: Prepare a list of clusters with corresponding mos for each;**
- **Step 2: Starting at the top of the list, calculate the cumulative mos and enter these figures in a column next to the measure of size for each unit;**
- **Step 3: Determine the range corresponding to each unit. The lower limit of the range is the previous row's upper limit of the range plus one. The upper limit is the – cumulative mos. Record the range in the column to the right of mos;**
- **Step 4: Calculate the sampling interval (SI) by dividing the total cumulative mos, by the number of units to be selected , that is:**
  - **SI = mos/number of unit to be selected**

Procedures for PPS sampling (2)

- **Step 5: Select a random number (r) between 1 and (SI). The r may be a decimal number. If it is, it is a good practice to retain 2 digits after the decimal;**
- **Step 6: Compute subsequent random numbers. These are obtained by adding the sampling interval (SI) to r; that is:**

$$r;\ r + SI;\ r + 2SI;\ \ldots;\ r + (a-1)SI$$

- **Step 7: Select each cluster whose range contains the number obtained in step 6;**

Note: In selecting clusters, the SI decimal points MUST be retained and the following rule is applied. When the decimal part of the sample selection number is less than 5, the lower numbered cluster is chosen, and when the decimal part of the sample selection number is 5 or greater, the higher numbered cluster is chosen.

Table 11.4.3: Selection of four Villages (clusters) using PPS

Cluster Num	mos: # HH listed	Cumulative mos	Range	Sample Selection #	Cluster Selected
1	163	163	1 - 163	22	*
2	250	413	164 - 413		
3	110	523	414 - 523		
4	210	733	524 - 733	638	*
5	207	940	734 - 940		
6	160	1100	941 - 1100		
7	165	1265	1101 - 1265	1254	*
8	180	1445	1266 - 1445		
9	140	1585	1446 - 1585		
10	309	1894	1586 - 1894	1870	*
11	245	2139	1895 - 2139		
12	325	2464	2140 - 2464		
Number of clusters to be selected = 4 ; SI = 2464 / 4 = 616; r = 22					
* = Selected clusters					

EXERCISE: PPS to select a sample of two villages (Part 1)

Village ID	Number of Households in Each Village	Cumulative mos	Range	Sample Selection Number
1	50			
2	30			
3	10			
4	20			
5	40			
6	30			

EXERCISE: PPS to select a sample of two villages (Part 2)

Village ID	Number of Households in Each Village	Cumulative mos	Range	Sample Selection Number
1	50	50	1 - 50	
2	30	80	51 - 80	*
3	10	90	81 - 90	
4	20	110	91 – 110	
5	40	150	111 - 150	*
6	30	180	151 - 180	

Advantage of systematic sampling at the second stage

- The selected households come from throughout the cluster, rather than being concentrated in one or two areas, which could happen by chance if a simple random sample is used;
- If necessary or desirable, the actual sampling can be done easily in the field with a minimum of training of the field staff;
- It allows for easy verification of fieldwork, walking a specified pattern and checking the selected SSUs.

Kish Grid for selecting one respondent from several eligible within a household

Tail # of household questionaire	Selected indinidual's line number according to number of eligible individuals in the household									
	1	2	3	4	5	6	7	8	9	10
0	1	1	3	4	4	4	7	7	9	1
1	1	2	1	1	5	5	1	8	1	2
2	1	1	2	2	1	6	2	1	2	3
3	1	2	3	3	2	1	3	2	3	4
4	1	1	1	4	3	2	4	3	4	5
5	1	2	2	1	4	3	5	4	5	6
6	1	1	3	2	5	4	6	5	6	7
7	1	2	1	3	1	5	7	6	7	8
8	1	1	2	4	2	6	5	7	8	9
9	1	2	3	3	3	6	6	8	9	10