



Food and Agriculture
Organization of the
United Nations



How to develop Socio-Economic surveys for baseline assessments

Planning & Sampling

Action Against Desertification programme

Workshop on socio-economic baseline assessment and M&E planning
Ouagadougou, Burkina Faso, 21st – 25th March 2016



INTRODUCTION

- Project evaluation can be **prospective** or **retrospective**
- Prospective evaluations begin before the project activities and are carried out throughout the project cycle
- Collecting baseline data is the crucial step to develop a prospective evaluation of the project
 - Baseline data are collected prior to the implementation of the project activities
 - Baseline data can be compared to the evaluation of the project outcomes during the project
- Comparing baseline data with the outputs of the evaluation process allows to identify the extent to which the project achieves its intended outcomes



OUTLINE OF THE PRESENTATION

- Planning the socio-economic baseline assessment
- Sampling strategies
- Preparing the questionnaire
- Implementing the household survey
- Data elaboration



PLANNING THE SOCIO-ECONOMIC BASELINE ASSESSMENT

- The assessment of the socio-economic baseline relies on
 - Secondary data already available before the project at the national and local level - may even be not fully reliable, but provides material to help understanding the local context
 - Direct surveys (usually focused on households)
- Key steps to plan the socio-economic baseline assessment are
 - Defining the sampling strategy of the household survey – crucial to allow representativeness and consistency of results
 - Identifying the relevant information to be collected – suited for calculating the desired indicators – and drafting the questionnaire
 - Hiring enumerators to administer the household survey and implement the interviews
 - Filling in a data matrix that, after check and clean, can be elaborated with multivariate and inferential statistics



SAMPLING STRATEGIES [1]

- Surveys cannot (in most cases) be performed on the whole population analysed, due to resources (time, money) and workload
- A sample is “a smaller (but hopefully representative) collection of units from a population used to determine truths about that population”
- First question to be answered is thus: what is the population of interest? (e.g. national population; population of the project area; population of the villages of the project area where some activities are to be implemented)
- You might sample the entire population when your population is very small OR you have extensive resources OR you don't expect a very high response



SAMPLING STRATEGIES [2]

- When you have to take a sample from the population, three issues should be considered very carefully, because they influence the representativeness of the sample
 - Sampling procedure
 - Sample size
 - Participation (response)
- After defining the population of concern, the following steps should be taken
 - Specifying a sampling method for selecting household out of the sampling frame
 - Determining the sample size
 - Implementing the sampling plan
 - Sampling and data collecting



SAMPLING STRATEGIES [3]

- Samples can be selected basing on probability (**probabilistic sampling**) or not (**non-probabilistic sampling**)
- **Probabilistic sampling** requires that every unit in the population has the same chance of being selected in the sample

SIMPLE RANDOM SAMPLE

A list of the household included in the population is available. Each household is given a number and a lottery system is applied to select the sample units

SYSTEMATIC RANDOM SAMPLE

Systematic sampling involves a random start and then proceeds with the selection of every kth element from then onwards, e.g. to select every 10th name from a list

CLUSTER SAMPLE

Multi-stage sampling technique. Sample areas (e.g. districts or villages) are randomly selected. Within the selected sample areas, observations are selected along one of the other sampling strategies



SAMPLING STRATEGIES [5]

➤ Example of systematic sampling (probabilistic)

- You have to sample the population of a city of 50,000 inhabitants, you want to get 500 people in your sample
- You have available from the council the list of the people living there, with names and address
- You randomly select one person out of the list. This is the first sample unit
- You sample every 100th persons after him/her; all the inhabitants have the same probability of being selected

➤ Example of cluster sampling (probabilistic)

- You have to sample the population of a region with 10 departments
- You randomly select 4 departments out of 10
- Within each selected department, you take the list of the villages in alphabetic order and you apply systematic sampling to select 1/3 of the villages
- Within each selected village, you take the list of the inhabitants in alphabetic order and you apply systematic sampling



SAMPLING STRATEGIES [4]

- With **non-probabilistic sampling**, some units of the population have no chance of selection
- Sample selection is based on assumptions regarding the population of interest, made by the researcher

QUOTA SAMPLE

The population is first segmented into sub-groups with different features. Judgment is then used to select units from each sub-group based on a specified proportions

CONVE- NIENCE SAMPLE

Also known as accidental sampling. The sample is drawn from that part of the population which is close to hand, i.e. readily available and convenient

PURPOSING SAMPLE

Researcher chooses the sample based on who they think would be appropriate for the study, e.g. directly appoints who to interview



SAMPLING STRATEGIES [6]

➤ Example of quota sampling (non- probabilistic)

- You want to sample 400 people living in a department where a reforestation project has been implemented
- You divide the population into three sub-groups: those who develop farming as main activity, those who develop farming as second or third activity, those who do not develop any farming activity
- For each sub-group, you sample fixed quota of people (e.g. 200+100+100) only in small villages (e.g. less than 100 households) because you consider that this is more consistent with the topic of your project
- People from larger villages have no probability of being sampled

➤ Example of convenience sampling (non- probabilistic)

- You want to sample 400 people living in a department where a reforestation project has been implemented
- You go to some villages you already know and you administer the questionnaire to the people you meet in the road



SAMPLING STRATEGIES [7]

➤ Example of purposive sampling (non- probabilistic)

- You want to sample 400 people living in a department where a reforestation project has been implemented
- You go to some villages you already know
- You sample the households living on your road, but only if farming is the main economic activity for them
- Your choice should be explained and consistent with the topic of your study



SAMPLING STRATEGIES [8]

- How to determine the **sample size**?
- The size of the sample depends on the size of the population
- The recommended sample size can be calculated through websites (e.g. <http://www.raosoft.com/samplesize.html>) providing an automatic calculation based on:
 - Population size
 - Desired confidence level, i.e. the amount of uncertainty you can tolerate in generalizing the results of the survey to the whole population (usually 95%; higher confidence level requires a larger sample size)
 - Margin of error, i.e. the amount of error that you can tolerate in the generalization of the results of the survey (usually 5%; lower margin of error requires a larger sample size)



SAMPLING STRATEGIES [8]

- How to determine the **sample size**?
- Example

Population size	Confidence level	Margin of error	Recommended sample size
50,000	95%	5%	384
50,000	99%	5%	655
50,000	95%	1%	8,507
10,000	95%	5%	370
10,000	99%	5%	623
10,000	95%	1%	4,900



SAMPLING STRATEGIES [9]

- Besides the sample you may also want to survey a “**control trial**” of the population
- A “control trial” is a (small) sample of the **population not targeted by the project**, i.e. not benefited by the project activities
- The survey is repeated on the control trial whenever done on the “real” sample
- A **counterfactual analysis** is performed between the data of the two samples, along the following question: *what would have been the value of the X indicator if the project had not been implemented?*
- The assessment of the outcomes of the project in the medium-long term is more solid and reliable



SAMPLING STRATEGY: THE GAMBIA CASE

- Probabilistic sampling
- Cluster sample (multi-stage sampling technique) → simple random sample
- Project population: ≈ 5000 HHs
- Sample size: 357 (95% confidence level, 5% margin of error)
- Census data: # of HHs disaggregated by cluster, district, and village in the project areas
- # of villages in each cluster and district determined proportionally to respective # of units
- # of HHs in each village, district and cluster determined proportionally to respective # of units and sample size
- Villages and HHs selected from lists with random # generator





Food and Agriculture
Organization of the
United Nations



Thanks for your attention !

