NITI Aayog

DMEO
DEVELOPMENT MONITORING AND EVALUATION OFFICE

SAMPLING GUIDELINES

सत्यमेव जयते

NITI Aayog

SAMPLING GUIDELINES

February 2022

Analysis by DMEO, NITI Aayog

Authors are Venugopal Mothkoor, Director, DMEO, NITI Aayog, Krishn Kant Sharma, M&E Lead, DMEO, NITI Aayog

www.dmeo.gov.in
www.niti.gov.in

ACKNOWLEDGEMENTS

This document on Sampling Guidelines has been developed as a result of arduous efforts of DMEO, NITI Aayog along with a wide range of stakeholders across the Monitoring and Evaluation ecosystem. First of all, I would like to express my deep gratitude to distinguished experts in the field who provided valuable comments and inputs on the document. I would like to thank Dr. Sonalde Desai, Director, NCAER and Dr Santanu Pramanik, Deputy Director, NCAER for their inputs regarding the use-cases and relevant content in the document. I would also like to thank the team from UNICEF including Dr Maaike Bijker, M&E Specialist, Dr Vishal Shastri, Research and Evaluation Advisor and Dr Swati Saluja, Research and Evaluation Consultant for their inputs specifically on qualitative aspects in sampling and the difference between unit of analysis and observation. I would also like to thank J-PAL team led by Dr Harini Kannan, Research Scientist for their suggestions on simplifying the document for laymen. I would also express my gratitude for Dr G C Manna, Ex- Dg, NSSO and CSO, Govt. of India for providing his inputs based on large sample surveys conducted by Government.

The document has also benefitted from the inputs received from Mr. S P Srivastava, Director, Development Monitoring & Evaluation Office, NITI Aayog and all other colleagues at DMEO. Last but not the least, I would like to thank the team - Mr. Venugopal Mothkoor, Mr. Krishn Kant Sharma, Ms. Fatima Mumtaz, Ms. Sonal Garg, Ms. Maitreyee Purohit and Mr. Vijender Kumar for their outstanding efforts and dedication towards producing this document on Sampling Guidelines which would help the users in Monitoring and Evaluation space to understand and formulate sample design for their studies.

TABLE OF CONTENTS

1

INTRODUCTION

Governments need data to monitor and evaluation of poverty, employment, housing conditions, health and learning outcomes and other dimensions of living standards. Some of this data for monitoring and evaluation of public policy programmes across the world come from surveys of households, firms, institutions and other market players. The higher reliance on the survey is because it is both impractical and costly to measure all the population or what is otherwise called a census. In such a context, sampling provides a mechanism through which observations can be made on the larger group without having to examine the population entirely. According to World Bank, sampling is the process of randomly selecting units from the population of interest to represent characteristics of that population (World Bank, 2021).

The purpose of this document is to assist various govt and non-government stakeholders by providing basic foundation-level knowledge in designing a sampling framework and sampling instruments. Users of this document will be able to:

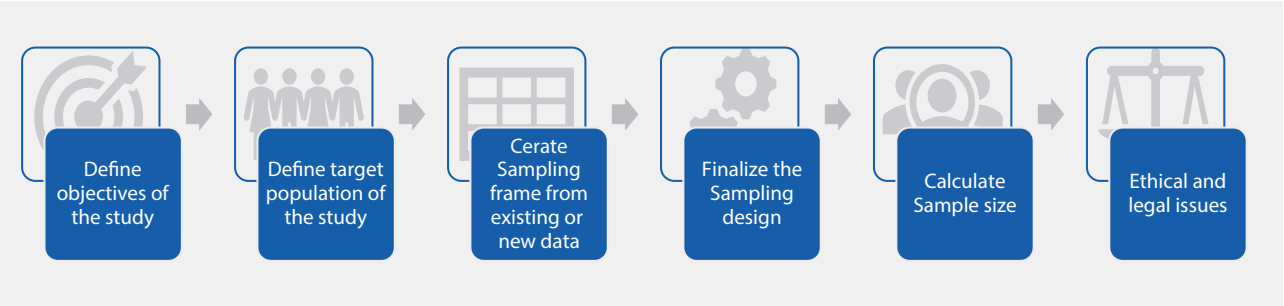
- Understand key concepts and definitions, methods, and steps of developing sampling design
- Identify ways to develop various sampling designs
- Select appropriate sampling design for implementing the survey

2

SIX STEP GUIDELINES

This document introduces six step general guidelines for designing a sampling strategy for any survey that needs to be conducted.

A. Objectives of the study
B. Target population
C. Sampling frame
D. Sampling design
E. Sample size
F. Ethical and legal issues



Surveys Conducted in India

National Statistical Office (NSO), is an apex survey institution in the country under the Ministry of Statistics and Program Implementation which is entrusted with the task of conducting large scale sample surveys across India. The data collected by NSO include household data, enterprise data, village facilities and land & livestock holdings (Seddey). The recent survey released by NSO include:

- Socio-economic survey-75th Round (2017-18): Education and Health Schedule
- Socio-economic survey-76th Round (2018): Survey of Persons with Disabilities schedule

- Socio-economic survey-76th Round (2018): Drinking Water, Sanitation, Hygiene and Housing condition

Other institutions also conduct surveys across India. For example, National Family Health Survey (NFHS) is an all-India survey conducted by the International Institute for Population Sciences (IIPS) in collaboration with a number of field organisations for the Ministry of Health and Family Welfare. The data collected predominantly fall under the subjects of family welfare and the health of the population. Till date, over 5 rounds of NFHS are executed with the first round in 1992-93, the second round in 1998-99, third round in 2005-06, the fourth round in 2015-16 and the fifth round in 2019-20. The other popular all India surveys executed in India include the Indian Human Development Survey (IHDS) by the National Council of Applied Economic and Research in collaboration with the University of Maryland. Besides such pan-India surveys, various government departments at the state level also execute surveys to collect output and outcome information as a result of a program intervention by the respective authority.

A. UNDERSTANDING OBJECTIVES

The early stage of sample survey design should include a careful review of literature and engagement with experts. Such an approach will result in the determination of what approach works best, what hypothesis has been tested and which questions are best suited for the specific survey. The most important factor that determines the design of the proposed survey is basically the question of what the survey aims to address. Easier than stated, this is the most complex task as the objectives of the survey vary from each agency involved in the process. Also, a survey does not exist in a vacuum. To a large extent, there exist many other surveys and the topic covered in them may need not be part of the new survey.

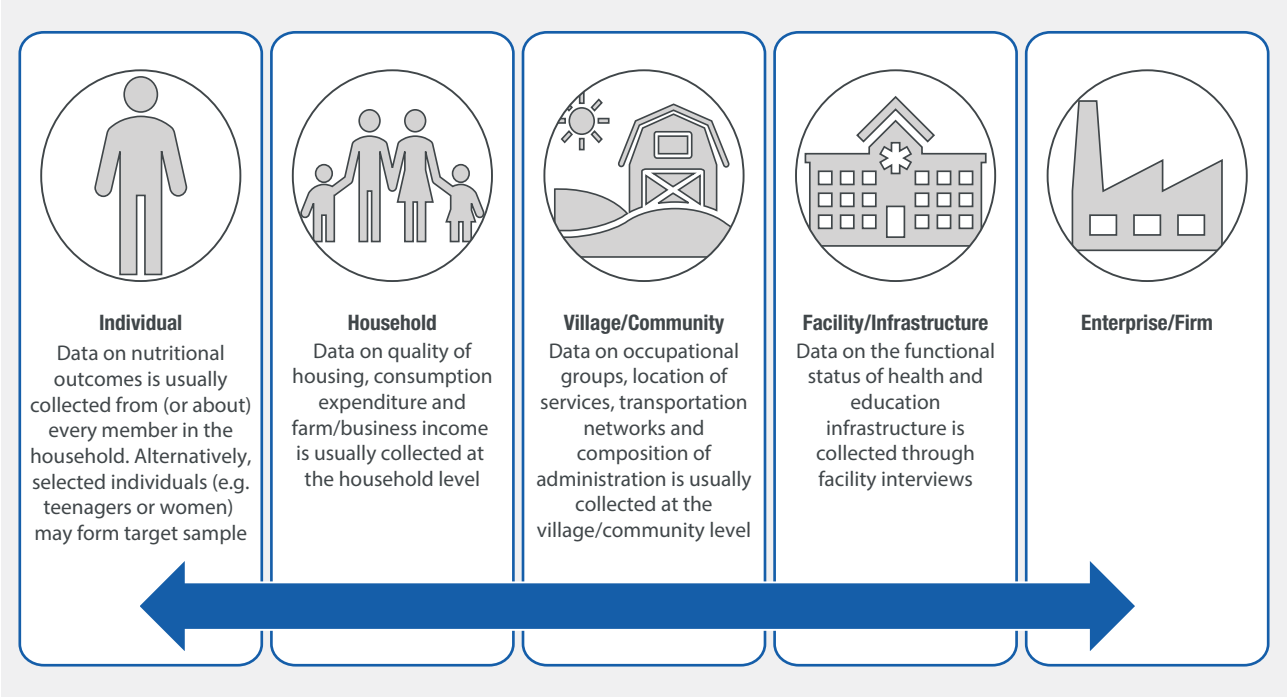
A.1 Single vs. multi-topic survey: The first question to enquire is whether there is a need for a single topic survey or multi-topic survey1. The various factors that may decide this choice between single vs. multi-topic surveys include the length of the questionnaire, funding availability, institutional capability and coherence with project objectives. Further, the length of the questionnaire also critically influences the quality of data being collected due to respondent and interviewer fatigue. The experience indicates that a significant percentage of the survey cost goes to implementation with travel and personnel cost being the two most important cost categories (Larossi, 2006). The field experience dictates that personnel cost needs to be suitably aligned to the nature of work and qualifications otherwise it may source resentment as a result of which survey quality suffers.

A.2 Data to be collected: The nature of the research dictates the type of data collected from the survey. For example, an impact evaluation study will typically require impact, outcomes, and outputs data and the factors that influence these output/outcomes. A descriptive study may require collection of only relevant indicators.

1 Single topic survey refers to survey administration pertaining to a single subject like drinking water while a multi-topical survey covers a wide range of subjects through a single survey on the field such as a collection of data on education, health and household assets etc.

B. TARGET POPULATION

The nature of research will also dictate at *what level* the data needs to be collected. For example, a government intervention to improve status of health infrastructure not only involves collection of data at health facility level but also at household/individual level to see whether the improvement in health infrastructure resulted in better health outcomes at the household/individual level. Five most common levels of observation used in evaluation studies include (World Bank, 2007):



It is important to ensure the eligibility of units for inclusion in the sample before recruiting them. Interviewers will need to be trained and supervised properly in this process so as to avoid sampling bias. The nature of the research design will also dictate the *frequency of data collection*. Robust evaluation design requires data at multiple time periods across multiple population-of-interest groups, often before and after the implementation of policies of interest.

Also, one needs to understand the key difference between a unit of observation and a unit of analysis. The unit of analysis is largely dictated by the research question while a unit of observation is the item you observe, measure or collect to say something about your unit of analysis. *For example, if one wants to know the dominant blood of various religious groups, the blood group is the unit of observation while a unit of analysis is the religious group. For one to compare dominant blood groups of different religious groups, these religious groups need to be adequately represented in the sample.*

Data collection through survey however should adhere to certain key principles outlined in BOSS strategy i.e., questions need to be **B**rief, **O**bjective, **S**imple and **S**pecific (Larossi, 2006).

As a rule of thumb, a question should not exceed 20 words and not have more than three commas (Larossi, 2006). There is need for avoiding leading questions, loaded questions and also be cautious of built-in assumptions while framing questions. Questions should be directly linked to what we intend to measure and due care must be taken to avoid redundancy.

After clearly identifying the research type and its objectives, the next step is to define the target population. The specification of the target population should be clear and complete so that all elements within the population are represented (National Audit Office, 2001). In practice, the actual population for the survey is smaller than the population forming the universe. *For instance, a labour force survey may include the working-age population in the age group of 15-59 years.*

Also, at times target population may be segregated into primary, secondary or tertiary purposes for certain types of evaluation. *For example, a survey among pregnant women may have pregnant women or women in reproductive age groups as the primary respondents. Her husband or mother-in-law may be secondary population groups as these are key influencers at the household level who affect decision making. Similarly, any community or village level influencer or someone from the health system could be a tertiary level population group.*

C. SAMPLING FRAME

The next most important step is to define the sampling frame on the basis of which the sample will be drawn. The sampling frame is the listing of the target population at different stages of sampling, including various characteristics of the sampling units. Very often, existing datasets such as census, pay-roll, government registries, voter lists can be relied upon as sampling frames. A sampling frame can also be geographical e.g., a list of pin codes. *A perfect sampling frame is one that is complete, accurate and up-to-date* (Department of Economic and Social Affairs, 2008). Any deviations in the compilation of complete and accurate sampling frame result in coverage error.

For example, recent surveys conducted by National Sample Survey (NSS) constructed separate sample frame for rural and urban areas based on the 2011 population census (NSSO, 2017). *For rural areas, census villages constituted the sampling frame while for urban areas, list of Urban Frame Survey (UFS) blocks constitute the sampling frame. It is important that variety of UFS blocks are covered given the large variation in urban areas and socio-economic clustering. However, one has to make sure that same areas are not selected in both rural and urban frame as frame for rural is census villages while frame for urban is latest updated UFS blocks.*

An example which explains the results when a sample frame is not accurate is the results of US presidential elections in 2016. American Association of Public Opinion Research set-up a committee to examine the reason why pollsters failed in predicting the outcome of US presidential elections. One of the reason the committee found is non-adjustment of over-representation of college graduates in the surveys as a result of which support for Clinton is over-estimated. Also, there are high number of voters who decided their presidential choice in the final week and the pollsters did not adjust weights accounting for them.

An example of non-coverage is where certain households living in temporary habitations are not covered by the survey as a result of which survey estimates are not truly representative of the population.

There is no need to be disheartened if there is no ready list available as the sampling frame. One can conduct an enumeration exercise to construct a sampling frame (Lundberg, 2010). However, experience indicates that in the cases where the household listing was costly and improbable due to security concerns, researchers relied on alternative techniques such as the use of satellite maps and area-based sampling (Himelein, Eckman, Murray, & Bauer, 2016).

Another aspect of the population that needs to be considered while developing a sampling frame is the nature of the population – its size, heterogeneity/homogeneity, accessibility, and spatial distribution. These factors play a critical role in the choices of sampling design. Larger sample size is required to study a heterogeneous population and non-response rate. It is also important to note that in a multi-stage design (discussed in sections below), the frame for each stage must be regarded as a separate component.

D. SAMPLING DESIGN

Most sample designs in developed and developing countries are multi-stage, stratified and clustered (Department of Economic and Social Affairs, 2005). Common guidelines on sample design structure include:

D.1 Why multi-stage? Multi-stage sampling is a technique in which sampling is done sequentially across two or more hierarchical levels. There are many reasons for the preference of multi-stage. First, the absence or poor quality of information on the target population makes it necessary to first select a sample of geographical units and then select a sample from the selected geographical units. Second, the use of multi-stage also controls the cost of data collection (Yansaneh).

For example, the NSSO survey uses a two-stage sampling technique. At the first stage, Census villages in rural areas and blocks in urban areas are selected as First Stage Units (FSUs). In the second stage, among the first stage selected units, households are selected.

Multi-stage sampling thus helps in the efficient selection of households from the pin-pointed locations (Department of Economic and Social Affairs, 2008). If the population is a large, widely scattered one, a single-stage sample (SRS) would be thinly spread, whereas a multistage sample concentrates the interviews in a number of locations.

D.2 Why stratified? Stratification helps in ensuring that the selected sample is spread across geographical areas and population sub-groups (Department of Economic and Social Affairs, 2008). The key advantage of stratification is that one is able to sample certain sub-groups at a higher sampling rate e.g. if we know that certain groups have more heterogeneity in outcomes. Stratification also helps counter the chance that one can land with a 'bad' sample where you don't get the sub-groups you want. Sample elements are independently selected from each stratum in a manner consistent with the objectives of the survey. Two important rules of stratification include (Department of Economic and Social Affairs, 2005):

- At least one sample unit must be sampled from each stratum. Further, strata are essentially independent and mutually exclusive subsets of the population; every element of a population must be in one and only one stratum.

- The second rule for stratification is that each stratum created should, ideally, be as different as possible. Heterogeneity among strata with homogeneity within strata be central feature guiding the establishment of strata.

For example, NFHS-3 uses the 2001 census list of villages as a sample frame which is then further stratified by a number of variables. The first level of stratification is geographic where districts are divided into regions. Within the region, villages are stratified based on variables such as village size, percentage of males in non-agriculture sector, percentage of population belonging to SC/ST and female literacy (International Institute for Population Sciences (IIPS), 2005-06).

Stratification ensures both representativeness and also provides a mechanism to estimate effect sizes at various stratified group levels provided the adequacy of sample sizes at that level. *For example, NFHS estimates in India are representative of the male and female population in various states.*

D.3 Why Clustering? Clusters are geographically defined units. Households in the same cluster generally tend to be more alike in terms of survey characteristics than households in general. Clustering reduces the cost of data collection considerably. However, clustering decreases the reliability of the sample due to the likelihood that people living in the same cluster tend to be homogenous i.e., Essentially it reduces the effective sample size). The so-called clustering effect sometimes needs to be compensated by increasing sample size (Department of Economic and Social Affairs, 2005). The difference between stratified and cluster sampling include:

- In stratification, all strata are represented in the sample. However, in clustering, a selection of the cluster itself is made.
- Strata are created to be internally homogeneous but external heterogeneous with respect to survey variables to be measured. The opposite is true for clusters.

The justification for cluster sampling is the economy it creates for sampling and data collection. The sampling economies are considerable with cluster sampling

- Listing needs to be done only in the selected final stage clusters (village/ urban wards)
- With data collection by face-to-face interviewing, multistage cluster sampling can give substantial savings in interviewer travel cost

The clustering or stratification effect can be measured by calculation of design effect (*deff*), which explains how large the sample variance is for clustered and stratified samples compared to Simple Random Sampling (SRS) of the same size. The design effect is dependent on two factors: i) intra-cluster correlation: it describes how strongly units in the same group relate with each other ii) Number of clusters. On the central question of what is the recommended design effect, information as available from previous national surveys of a similar design to the proposed survey may be relied upon. Principle for low *deff* include (Department of Economic and Social Affairs, 2005):

- Use as many clusters as feasible. There is a trade-off between cost and efficiency. The choice of a large number of clusters also means higher costs.
- Use the smallest cluster size in terms of the number of households that is feasible. Also, one should be mindful of the fact that estimation of certain contextual factors within a cluster, one needs a reasonable sample size for the cluster.
- Use constant cluster size than varying.

A stratified design more complex than SRS has the potential to be more efficient than SRS, if strata means are different with respect to the outcome of interest. The greater the heterogeneity in strata means, more will be the gain in precision relative to SRS. Hence, it is important to choose the stratifying variables carefully. It leads to a more representative sample as the sample covers all strata

Sample size needs to be adjusted for design effect. For example, a sample survey size of 5000 needs to increase to 7500 (5000*1.5) by accounting for a design effect of 1.5.

D.4 Selection methodology of First Stage Units (FSUs): The most popular used in the selection of FSUs is the probability proportional to size (PPS) method. PPS ensures constant selection probability at the final stage since each cluster (irrespective of size) has a fixed number of units sampled. (See UNICEF document for more details). In real life, census villages or wards in urban areas are of varying sizes. PPS takes this into consideration wherein large clusters have a greater probability of selection while smaller clusters have a lower probability (WHO).

First stage: PPS sampling at first stage mean, large clusters have high probability of being sampled

Second stage: Since exactly same number of units per cluster need to be sampled, individuals in large cluster have smaller probability of being sampled

Overall: Second stage compensates first stage, so that each individual in the population has the same probability of selection.

For example, the selection of FSUs by NSSO also uses the PPS technique. For the rural sector, from each stratum/sub-stratum, the required number of sample villages are selected using PPSWR (Probability Proportional to Size With Replacement) technique with size being the population of the village as per census 2011. Similarly, for the urban sector, FSU is selected using PPSWR with size being the number of households in the UFS block. Also, the total number of sample FSU allocated to each state/UT is proportion to population as per census 2011 subject to a minimum sample allocation to each state/UT.

PPS technique selection also varies with the type of survey. Farm survey uses the size of a farm as a measure for selection as typical farm parameters such as income, crop production and expenses are correlated with farm size. Similarly, a business survey uses measures of size which include no. of employees, turnover or sales. The selection of FSUs

also needs to consider the required heterogeneity in the FSU population. Estimation of benefits on various socio-economic groups necessitates the selection of an FSU unit where these groups are adequately represented.

D.5 Selection methodology for Ultimate Stage Units (USUs): Once the sample selection of FSUs is completed, often a listing exercise is carried out in the selected FSU with the objective to create an up-to-date sample frame from which households can be selected. Based on the availability of the list, various *probability sampling techniques* described in Annexure-1 to select the households may be employed.

In contrast to probability sampling, there exist *non-probability sampling methods* to guide the selection of sample units. However, there exists no statistical theory to guide the use of non-probability selection. Often due to the concerns of cost and convenience, there is a switch to use non-probability samples. The various non-probability sample selection tools are described in Annexure-2 (Box-1 introduces the reader to key differences between probability and non-probability sampling).

Typically, when FSU are small < 300 USUs or so, all USUs are listed and a subset is selected. When FSUs are large, the FSU is divided into several segments and 2-3 segments are fully listed. Selection bias may be introduced if outlying hamlets are omitted from the listing exercise. In actual practice, the probability sampling methods described are employed in survey design. For example, NSS and NFHS use probability sampling techniques. One should also be mindful of the fact that listing in USUs is the hardest part of sampling.

Box-1: Key differences between probability and non-probability sampling

Probability sampling is a scientific approach of sample selection that satisfies certain conditions:

- Each element in the population has a (known) nonzero probability of selection
- A sample is selected by a random mechanism (implemented by some type of algorithm)

Advantages of probability sampling

- Selection biases are avoided
- Randomly selected samples are viewed as objective
- Statistical theory can be used to derive properties of the survey estimators

Non-probability sampling includes convenience sampling, judgement or purposive sampling, quota sampling. For example, volunteer subjects for studies; the patients of a given doctor; persons posting their comment on Facebook

Main weaknesses of nonprobability sampling

- Subjectivity, bias
- Precludes the development of a theoretical framework for it
- Can be assessed only by subjective evaluation

Non-probability sampling is still widely used in practice, mainly for reasons of cost and convenience (e.g., in the absence of sampling frame). Subsequent sections in the report discusses the various types of probability and non-probability sampling.

E. SAMPLE SIZE

The sample size is dependent on factors such as: i) Margin of error or Precision ii) Amount of variability iii) Confidence level iv) Population size v) Population proportion (National Audit Office, 2001).

- ***The margin of error*** describes how close we can expect a survey result to fall relative to the true population value. The law of diminishing returns underlies the relation between sample size and margin of error. Other things being equal, the error is inversely proportional to the square root of the sample size. The smaller the margin of error, the larger is the sample size needed. One can also use an online calculator based on the availability of information on population size, sample size and confidence level (See https://aytm.com/pages/mes).
- ***The amount of variability*** in the population will also affect the accuracy and therefore the size of the sample. The higher the variability in the sample, the larger is the sample size required.
- The ***confidence level*** is the likelihood that results from the sample lie within the associated precision. The confidence level describes the level of uncertainty in the sample mean or prevalence as an estimate of the population mean. 95 percent CI is almost universally taken as the standard and the sample size necessary to achieve it is calculated accordingly (A critical value of 1.96 is associated with 95% CI). The sample size increases as the degree of statistical confidence increases.
- ***Population size*** does not normally affect sample size. In fact, the larger the population size, the lower the proportion of that population that needs to be sampled to be representative provided the population is homogenous. However, a larger heterogenous population may require more sample size. It is only when the proposed sample size is only a small percentage of the population, then population size becomes important to calculate the sample size.
- ***Population proportion*** is the proportion of items in the population displaying the attributes that one is seeking. Population proportion in other words is the estimated baseline value of the indicators one wants to measure through the survey. Values closest to 50% are the most conservative, requiring the largest sample size.

For evaluation studies, determination of a programmatically significant effect of the intervention is crucial and this impacts sample size besides factors discussed above. The minimum effect of intervention is often more important than the margin of error especially in surveys conducted for evaluation. Though a larger sample size may help detect even small changes in the intervention, one needs to carefully balance it against the cost considerations.

One may use online tools available for sample size estimation by specifying parameters of margin of error, confidence level, population size and population proportion. For example:

- http://www.raosoft.com/samplesize.html
- https://www.abs.gov.au/websitedbs/d3310114.nsf/home/sample+size+calculator

One can also use statistical tools such as stata for sample size estimation. See https://dimewiki.worldbank.org/Power_Calculations_in_Stata

The big challenge while designing a survey is to analyse many characteristics of the population. This is practically possible by the use of *stratification*. For example, drawing representative estimates at the state level and industrial sector level requires the use of state and sector as a key stratification variable. *Sample sizes are then are to be drawn at each individual strata level* and summed to determine the sample size for the whole survey. However, cost and availability constraints may force sometimes to cut the sample size. Sample size also needs to be adjusted for non-response. The reasons for non-response may include: i) No one at home ii) Vacant unit iii) Demolished or uninhabitable unit iv) Refusal v) Away temporarily (Department of Economic and Social Affairs, 2005). This may necessitate rationalization of strata. JPAL suggested six key steps for finalizing sample size (JPAL, 2018):

- A larger sample increases the statistical power of the evaluation
- If the effect size of the program is small, evaluation needs a larger sample to achieve a given level of power
- An evaluation of a program with low take-up needs a larger sample
- If the underlying population has high variation in outcomes, the evaluation needs a larger sample
- For a given sample size, power is maximized when the sample is split equally split between treatment and control.
- For a given sample size, randomizing at cluster level instead of individual level reduces the power of evaluation.

E.1 Non-sampling error: Non-sampling error is error associated with data collection and processing procedure. This may arise due to invalid definitions and concepts, inaccurate sample frames, unsatisfactory questionnaires, defective methods of data collection, tabulation and poor coverage. Often non-sampling errors are not easily controllable and are unpredictable. This error increases with sample size and can be more damaging (Department of Economic and Social Affairs, 2005). UNSTATS classifies non-sampling error into:

- Specification error: Occurs when the concept implied by the question is different from the underlying construct that should be measured
- Coverage error: Occurs when certain units are omitted or duplicated due to imperfect demarcation of selected sampling areas. This may occur when old and non-updated sampling frames are relied for selection.
- Non-response error: Occurs due to failure to obtain responses from the sample unit.
- Measurement error: Occurs when measured value departs from the actual value.
- Processing error: Comprising of editing errors, coding errors, data entry errors and programming errors.

Some of the mechanisms through which non-sampling errors can be minimized include i) Consistency check ii) Sample check iii) Reinterview checks iv) Quality control which includes interviewer training and question design that provides for easy recall for respondents (Department of Economic and Social Affairs, 2005).

E.2 Sample Weights: Sample survey are susceptible to various issues of non-coverage, non-response and selection of units with unequal probabilities. In such scenario, sample weights become essential to correct for such issues and estimate parameters of interest. The estimates especially at the population level need to be weighed in-case of samples not self-weighted (Larossi, 2006). Two most common types of weights (Johnson, 2008):

- Design weights: Normally used to compensate for over or under-sampling of cases.
- Post stratification or non-response weight: Used to compensate for the fact that person with certain characteristics is less likely to respond. It is estimated based on ratio of proportion of certain strata of population in the census to that of the sample.

Total weight that is applied is usually the product of both design and post-stratification weight (Department of Economic and Social Affairs, 2005). Imbalance in weights also result in an inflation of standard errors – this is another area where advance planning can help e.g., good designs will avoid high imbalance between strata so that weight ranges are not too wide.

F. ETHICAL AND LEGAL ISSUES

There should be a balance between the ethical and legal requirements and the sampling designs. Concerns related to informed consent, privacy, confidentiality, anonymity, etc., affect the sampling design and data collection. For example, researchers may set up legal and ethical cells to address any untoward issue during the data collection on gender-based violence. If any respondent reports violence during the data collection, the research team may report to the suitable authority and arrange counselling as per the legal and ethical research protocols.

Key ethical considerations in sampling design should consider:

- Do not let sampling strategy inadvertently omit key groups whose experiences may be unique (e.g., women, Scheduled Caste hamlets)
- Any restrictions placed in sample selection should have strong substantive justification (e.g., omitting ages 60 and above from labour force surveys).
- Informed consent for initial interview and recontact (if undertaking repeat interviews)

3

BEST PRACTICES IN SURVEY DESIGN SPECIFICATIONS

According to Samsha, the best practices in survey design include (OFFICE OF MANAGEMENT AND BUDGET, 2006):

- Include following in survey design: i) Target population ii) Response rate goals iii) Frequency and timing of data collection iv) Data collection methods v) Sample design vi) Sample size based on precision requirements
- Sample Design to include elements such as: i) Sample Frame ii) Sampling unit iii) Sampling strata iv) Size by each stratum v) Probability of selection of unit vi) Estimation and weighting plan
- According to UNSTATS, the characteristics of a good sampling unit include: i) Presence of clear identifiable boundary ii) Adequate coverage of target population iii) Large in number and iv) Data available for stratification purposes. If boundaries of the sampling unit are unclear then larger and more clearly defined units need to be used as a sampling unit. Also, small sampling units may be combined while large sampling units may be segmented to satisfy the requirements of sample design (Yansaneh).
- Discuss the reasons for the selection of non-probability sampling method if any employed and methods incorporated to reduce the bias inherent in such methods.
- Confidentiality pledge
- Data collection plan to include: i) Methods of data collection for achieving acceptable response rates ii) training plan iii) Data analysis plan
- Adjust sample estimates for sampling error and non-sampling error

ANNEXURE-1: PROBABILITY SAMPLING METHODS

Simple Random Sampling: A simple random sample is utilized when all the units in the population have an equal probability of being included in the sample. Each individual is chosen entirely by chance. The tool available at https://www.randomizer.org/ can be used for generating a list of random numbers. The random numbers then are matched to the dataset to select the household members who need to be interviewed.

Systematic Random Sampling: Systematic Sampling is a method in which a sample is obtained by selecting every k^{th} element from the population where k is an integer greater than 1. Use the excel tool available at the below location for finalizing the list of households to be interviewed.

https://www.aphis.usda.gov/animal_health/CEAH_toolbox/Protectedsamplesizecalculators/Protectedsamplesizecalculators/html_interval_sampling_calculator.xls

However, if there is a periodic variation in the population, systematic sampling can yield results that are either under-estimates or over-estimates. Also, the selection method is prone to abuse by enumerators (Department of Economic and Social Affairs, 2005).

Stratified Random Sampling: Stratified sampling is a method in which sampling units are divided into a heterogeneous group that are internally homogenous. Within each stratum, sampling and estimation procedures can be different or similar. The main advantage of stratified sampling is the increase in the precision of estimates.

Cluster Sampling: Cluster sampling is a method wherein sampling units are a group of units to be studied. The key to clustering is to divide the population into the discrete group in such a way that there is no difference between the groups in relation to what we want to measure. This means variability that exists in the population is contained within the cluster. As a result, we can select a few of these clusters randomly to conduct the study. Clustering is preferred as it minimizes the cost of the sampling process. However, there is a trade-off which is loss of precision since there is a lack of heterogeneity among the clusters.

These probability sampling methods such as simple random sampling, systematic random sampling, stratified sampling and cluster sampling can also be applied in the selection of FSUs.

ANNEXURE-2: NON-PROBABILITY SAMPLING METHODS

Judgment Sampling: A method wherein experts choose the sample elements. The main justification for advocacy of this method is that random technique may lead to the selection of bad samples. Ironically, it is highly dependent on the choice of experts themselves.

Random Walk Sampling: A method in which enumerators spin a bottle to determine a random direction in which they walk and select every k^{th} household (World Bank, 2021). Another method of random sampling is the selection of a random starting point and the direction of the interviewer's position is assigned at 90-degree intervals. Intuitively, a random walk would only be unbiased if paths taken during selection crossed each household once and only once, which is unlikely in the field. Several studies show that data collected on random walk do not match the population on basic demographics such as age, sex, education, household size and marital status (Himelein, Eckman, Murray, & Bauer, 2016).

Quota Sampling: Quota sampling divides the survey population into mutually exclusive sub-groups based on non-random features, traits or interests. Within each sub-group, interviewer often employs random walk or other non-probability sampling techniques to select household samples until the quota is reached.

Convenience Sampling: Convenience sampling is a method wherein people who are easy to reach are selected in the sample. The obvious advantage of the method is easy to use but the advantage is greatly offset by presence of bias.

BIBLIOGRAPHY

World Bank. (2021, Dec 29). *Sampling*. Retrieved from DIME Wiki: https://dimewiki.worldbank.org/Sampling

Seddey, S. C. (n.d.). AN OVERVIEW OF NATIONAL SAMPLE SURVEY ORGANISATION (NSSO). Retrieved from https://users.pop.umn.edu/~rmccaa/ipums-global/india_nsso_durban_workshop.pdf

Larossi, G. (2006). *The Power of Survey Design : A User's Guide for Managing Surveys, Interpreting Results, and Influencing Respondents.* Washington, DC: World Bank.

World Bank. (2007, Oct). *Data for Impact Evaluation.* Retrieved from https://documents1.worldbank.org/curated/zh/332891468313760995/pdf/423800NWP0Doin10Box321452B01PUBLIC1.pdf

Department of Economic and Social Affairs. (2008). *Designing Household Survey Samples: Practical Guidelines*. Retrieved from https://unstats.un.org/unsd/demographic/sources/surveys/Series_F98en.pdf

National Audit Office. (2001). *A Practical Guide to Sampling.* Retrieved from https://www.nao.org.uk/wp-content/uploads/2001/06/SamplingGuide.pdf

NSSO. (2017). *Instructions to Field Staff.* Retrieved from https://www.mospi.gov.in/documents/213904/533217//Chapter%2011602843195585.pdf/d819dc82-bb5a-f552-28bc-68b07bd1f065

Lundberg, M. (2010). *Sampling for Evaluation.* Retrieved from https://thedocs.worldbank.org/en/doc/966121525895271435-0090022018/original/35.SamplingLundbergENG.pdf

Himelein, K., Eckman, S., Murray, S., & Bauer, J. (2016). *Second-Stage Sampling for Conflict Areas:Methods and Implications.* World Bank.

Department of Economic and Social Affairs. (2005). *Designing Household Survey Samples: Practical Guidelines.* Retrieved from https://unstats.un.org/unsd/demographic/sources/surveys/Handbook23June05.pdf

Yansaneh, I. S. (n.d.). *Overview of sample design issues for household surveys in developing and transition countries.* Retrieved from https://unstats.un.org/unsd/hhsurveys/pdf/Chapter_2.pdf

International Institute for Population Sciences (IIPS). (2005-06). *National Family Health Survey.* Retrieved from Metadata: https://microdata.worldbank.org/index.php/catalog/1406

WHO. (n.d.). *Steps in applying Probability Proportional to Size (PPS) and calculating Basic Probability Weights.* Retrieved from https://www.who.int/tb/advisory_bodies/impact_measurement_taskforce/meetings/prevalence_survey/psws_probability_prop_size_bierrenbach.pdf

JPAL. (2018). *Six rules of thumb for determining sample size and statistical power.* Retrieved from https://www.povertyactionlab.org/sites/default/files/research-resources/2018.03.21-Rules-of-Thumb-for-Sample-Size-and-Power_0.pdf

Johnson, D. R. (2008). *Using Weights in the Analysis of Survey Data.* Retrieved from https://pages.nyu.edu/jackson/design.of.social.research/Readings/Johnson%20-%20Introduction%20to%20survey%20weights%20(PRI%20version).pdf

OFFICE OF MANAGEMENT AND BUDGET. (2006). *STANDARDS AND GUIDELINES FOR STATISTICAL SURVEYS.* Retrieved from https://www.samhsa.gov/data/sites/default/files/standards_stat_surveys.pdf

ABOUT THIS DOCUMENT

The purpose of Sampling Guidelines document is to assist various govt and non-government stakeholders by providing basic foundation-level knowledge in designing a sampling framework and sampling instruments.

ABOUT DMEO, NITI AAYOG

The Development Monitoring and Evaluation Office (DMEO), attached to NITI Aayog, is the apex monitoring & evaluation (M&E) office in the country, with a mandate to drive evidence-based policy making through M&E of government policies and programmes. Since its inception in 2015, the Office aims to shift the discourse of public policy towards rigorous, data-driven, citizen-centric, and decentralized policymaking, to improve governance and facilitate the formation of a New India.

Visit us at www.dmeo.gov.in

सत्यमेव जयते

NITI Aayog

DEVELOPMENT MONITORING AND EVALUATION OFFICE

EXPLORE MORE

of DMEO's thought leadership articles by visiting https://dmeo.gov.in/articles

This document is produced by advisors and consultants at DMEO, NITI Aayog, as general guidance. It is not intended to provide specific advice. If you require advice or further details on any matters referred in this document, please contact DMEO.