The Nuts and Bolts of Citizens' Surveys

—Vivek Misra

The most powerful weapon on earth is public opinion - never forget that.
Paul Crouser

1 Introduction

In old times and new, 'Governance' has been the subject of much interest and debate. Be it the 'Arthashastra', the model Greek city-states, Marx's socialistic doctrine, the neo-liberal ideology or the recent Good Governance agenda – attempts have been made, time and again, to define and redefine relations between the ruler and the ruled. However, unlike earlier attempts, the latest one comes at a time when an information revolution is sweeping across the world. This holds the potential of fundamentally changing the relationship between the ruler and the ruled, or rather, the government and the citizens.

With increasing access to information and the democratisation of knowledge, the role of citizens in the process of governance is undergoing a paradigm shift. From being mere recipients of a one-way dialogue (or should we say monologue), citizens now increasingly demand to be consulted and participate in the decision-making process, rather than just serve as the `sleeping mass' in a representative democracy. In light of the changing aspirations of the citizens, it is incumbent on the State to provide ways and means to enhance the role of citizens in the decision-making process and facilitate better State-Society articulation.

Citizens' Surveys serve as an important tool by which the State/Government can engage citizens in the process of governance. Not only do citizens' surveys provide inputs that aid and enable the government to frame policies, evaluate programmes, assess and improve service delivery, map attitudes and preferences, study voting intentions and examine demographic/socio-economic profiles but also by definition, surveys constitute a two-way communication process that enhances the nature and quality of articulation between the government and the citizens. In other words, citizen surveys are `a good in itself and of itself'.

While customer surveys have found an important place in the management world, top-down prescriptions have historically rendered their application to the public sector a rare phenomenon. This has been typically true of countries which have lacked a vibrant civil society. However, in recent times, the downward pressures unleashed by the twin forces of democratisation and information revolution in a fractious and competitive national/international political order have necessitated the need for the Government to use citizens' surveys in order to engage with citizens.

The growing dependence on surveys and the relative lack of knowledge on survey methodology has lead to a significant demand for information on the subject. The main purpose of this paper is to supply this information in simple, non-technical language for use by Government Departments. An equally important purpose of this paper is to identify problems that may arise during development of a survey and to provide techniques and guidance for solving these problems.

The paper has four subsequent sections. The second section attempts to contextualise the role of citizens in the governance process and highlights how citizens' surveys are important within this context. The third section gives a brief description of research approaches. Given the fact that surveys fall within the broader ambit of social sciences research, it is imperative that a rudimentary understanding of basic research approaches to information gathering precede the understanding of survey research.

The fourth section is the main component of the paper. It is broadly divided into six subsections. The first subsection provides an introduction to survey research viz. the basic definitions, types of surveys and key steps involved in surveying. The second subsection details the first four steps involved in survey design namely, defining the purpose of the survey, developing hypothesis, defining the population/target segment and developing the survey plan. In the third subsection, the paper attempts to elaborate on sampling techniques and methodology. The next subsection looks at the survey instrument and suggests ways in which the effective questionnaires can be developed. The fifth subsection describes the final steps in surveying namely, data gathering, data reduction and data analysis.

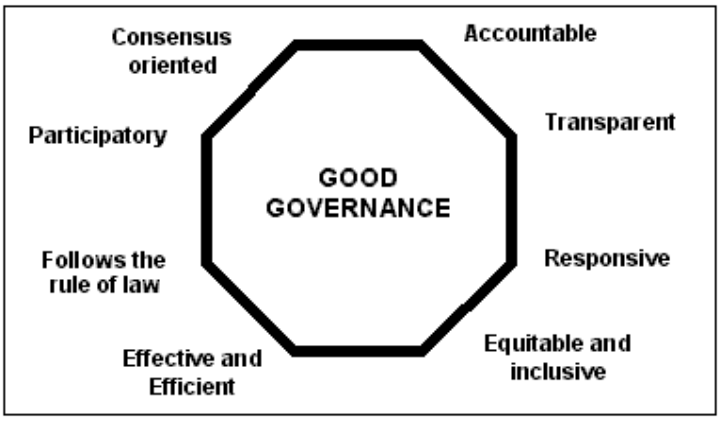
The fifth and final section of the paper provides a hypothetical case study that traces every stage in the survey process and presents a list of dos and don'ts that government departments/agencies should keep in mind while undertaking a survey exercise.

Surveys involve varied and complex procedures. This paper only highlights the major information, techniques, and procedures. It has been so constructed that only a rudimentary knowledge of statistics is required. More detailed treatments of these subjects can be made available on request. Although many of the techniques and procedures covered here apply equally well to different types of survey, the primary focus is on surveys through personal face-to-face interviews which are the most common form of surveys in India. Finally, this paper should not be perceived as a do-it-yourself kit on surveys. It provides relevant information for departments to understand the basics of survey methodology. For conducting robust and accurate citizens' surveys, it is advised that they employ the services of well known research agencies that have the skill and capacity to undertake such assignments.

2 Good Governance, the Role of Citizens and Citizens' Surveys

Governance can be defined as `the manner in which power is exercised in the management of a country's economic and social resources for development'. The nature of governance is reflected in two dynamic processes (a) the process of decision making and (b) the process by which decisions are implemented (or not implemented). Good governance is the effective implementation of policy and provision of services that are responsive to citizen needs.

Good governance has 8 major characteristics. It is participatory, consensus oriented, accountable, transparent, responsive, effective and efficient, equitable and inclusive and follows the rule of law. It assures that corruption is minimised, the views of minorities are taken into account and that the voices of the most vulnerable in society are heard in decision-making. It is also responsive to the present and future needs of society.



Characteristics of Good Governance

Citizen engagement is a core element of good governance

Government programmes and policies must be in a constant state of evolution in order to meet the public's changing needs and expectations. When a government allows its institutions to ossify, it is no longer serving the public good. Strengthening relations with citizens enables government to do just that. It allows government to tap new sources of policy-relevant ideas, information and resources when making decisions. Equally important, it contributes to building public trust in government, raising the quality of democracy and strengthening civic capacity.

To engage people effectively in policymaking, governments must invest adequate time and resources in building robust legal, policy and institutional frameworks. They must develop and use appropriate tools, ranging from traditional opinion polls and surveys of the population at large to consensus conferences with small groups of laypersons.

Citizens' Surveys assume importance in this context. Unlike in the private sector where the market mechanism and continuous customer surveys provide feedback to private sector managers, in the public sector feedback from the public comes from interest groups and squeaky wheels. Feedback from the bulk of the public or the silent majority comes only at election time and provides little guidance towards making service delivery more effective and efficient. Surveys can be used to close this feedback gap and to gauge the effectiveness of their operations, identify unmet public needs and improve service delivery.

Citizens' Surveys can enable governments in:

- Better allocation of resources: As fiscal pressures on governments have increased, setting priorities and allocating resources where they are most needed has become increasingly important. Commonly used approaches to balancing budgets may seem equitable, but in practice can shortchange citizens. Citizens may value some services more than others and they may be more satisfied with some than with others. When government managers compare the importance of various government services to the level of service satisfaction, they can create a powerful tool for making resource allocation

decisions. They can use such information as a basis for reallocating resources from services that citizens rank low in importance to those they rank higher.

- Optimising service levels: In addition to supporting resource allocation decisions and focusing management attention on unsatisfactory services, citizen surveys can be used to assess the levels of service governments should provide By making such investments, a government can ensure that it is not over- or under-providing the service and can more closely match level of service to citizen demands. Although surveys cannot be used to determine the precise level of response that should be provided, they can provide valuable information on overall service expectations.

- Performance Evaluation: Evaluating the performance of public-sector organisations is much more difficult. No single financial indicator can be used to distinguish high-performing governments from low performing governments. When evaluating performance, public sector managers tend to focus on the volume of resources used and activities performed. Incorporating the results of a citizen survey into the performance evaluation process can present a broader, more accurate view of government services delivery.

- Framing better policies and programmes: The government could also obtain inputs through citizens' surveys that could enable it to make better policy and programmes. Important policy decisions that have large public ramifications may require public support. Equally important are inputs obtained from citizens' surveys which provide information on the impact of policy reforms on the public. Survey could also provide information on effectiveness of government programmes and schemes targeted at particular sections of the population.

- Setting user fee levels: Citizens' surveys can also determine the level at which general fund resources should subsidise services from which many residents benefit. For some services, appropriate subsidy levels may be determined simply by counting the number of citizens who receive the service and dividing by the total population. For many services, however, this method may understate the overall demand. Many residents may want certain services provided even though they themselves do not directly benefit form them. Citizen surveys can be extremely useful in determining the overall level of demand for such services and the extent to which they should be supported by general fund revenues.

There are many other uses of surveys. They could be used to examine demographic and socio-economic profiles of different target segments. They could also enable an assessment of attitudes and behaviour of voters. Surveys could also take the form of opinion polls and assess the opinion of citizens on issues of public interest.

Citizens' surveys, if effectively used, can be an important tool for citizen engagement in the process of engagement. The following sections in the paper attempt to provide a basic understanding of the nuts and bolts of a citizens' survey that government departments may find useful in their quest for greater civic engagement in governance.

3 Research Approaches to Information Gathering

It is necessary, at the outset, to provide a brief idea about research approaches. This will facilitate a better understanding and appreciation of the subject of enquiry namely, Survey Research. There are two major types of research approaches: Qualitative Research and Quantitative Research.

Qualitative research is concerned with the opinions, experiences and feelings of individuals producing subjective data. It describes social phenomena as they occur naturally. There are four major types of qualitative research design:

- Phenomenology or study of a phenomena and describes something that exists as part of the world in which we live
- Ethnography is a methodology for descriptive studies of cultures and peoples
- Grounded theory leads to development of new theory through the collection and analysis of data about a phenomenon
- Case study research is used to describe an entity that forms a single unit such as a person, an organisation or an institution

Quantitative research depends on the ability to identify a set of variables. Data are used to develop concepts and theories that help us to understand the social world. Quantitative research is deductive in that it tests theories which have already been proposed. Quantitative research can be broadly classified into:

- Descriptive or Survey research which involves studying the preferences, attitudes, practices, concerns, or interests of some group of people
- Correlational research that attempts to determine whether, and to what degree, a relationship exists between two or more variables
- Causal-Comparative research seeks to discover a cause-effect relationship between two or more different programmes, methods, or groups
- Experimental research in a form of correlational research that resembles an experiment

The table below compares the two approaches - Qualitative and Quantitative - on key dimensions.

Qualitative Versus Quantitative Research

Comparison Dimension	Qualitative Research	Quantitative Research
Type of research	Exploratory	Descriptive or causal
Purpose	Generate hypotheses	Test hypotheses
Types of questions	Probing	Non-probing
Sample size	Small	Large
Information per Respondent	Much	Varies
Administration	Requires interviewer with special skills	Fewer special skills required
Type of analysis	Subjective, Interpretive	Statistical, summarisation

Hardware	Tape recorders, projection devices, videos, pictures, discussion guides	Questionnaires, computers, printouts
Training of the researcher	Psychology, sociology, social psychology, consumer behavior, marketing, marketing research	Statistics, decision models, decision support systems, computer programming, marketing, marketing research
Ability to replicate	Low	High

4 Survey Research

4.1 Introduction

What is a Survey

Webster defines a survey as "the action of ascertaining facts regarding conditions or the condition of something to provide exact information especially to persons responsible or interested" and as "a systematic collection and analysis of data on some aspect of an area or group." A survey, then, is a process and goes much beyond than the mere compiling of data. To yield relevant information, the data must be analysed, interpreted and evaluated.

Types of Surveys

Surveys can be divided into two general categories on the basis of their extensiveness. A complete survey is called a "census." It involves contacting the entire group you are interested in -- the total population or universe. The other category is more common; it is a sample survey. A sample is a representative part of a whole group (universe). Thus a sample survey involves examining only a portion of the total group in which one is interested, and from it, inferring information about the group as a whole.

By sampling only a small portion of a large population, it is possible to collect data in far less time than would be required to survey the entire group. The smaller amount of data gathered by sampling as opposed to surveying an entire population can mean large cost savings. Finally, a carefully selected sample may yield more accurate information than a less careful collection of data from the entire population. Oh the other hand, there are certain disadvantages of sampling. The main disadvantages stem from risk, lack of representativeness of the sample, and insufficient sample size, each of which can cause errors. Inattention to any of these potential flaws will invalidate the survey results.

Surveys can be classified by their method of data collection. Mail, telephone interview, and in-person interview surveys are the most common.

- *Mail surveys* can be relatively low in cost. The main problems, however, with this type of survey are (a) the non-response errors associated with it and (b) lack of control on the representativeness of the sample that responds.

- *Telephone interviews* are an efficient method of collecting some types of data. They are particularly suited in situations where timeliness is a factor and the length of the survey is limited. However, the sampling frame in this kind of

survey may be much smaller than the actual universe, especially in the Indian context where access to basic telephone services is limited. Further, they may not be well suited in situations where detailed information may be required.

- Internet surveys: A more recent innovation in survey technology is the Internet survey in which potential respondents are contacted and their responses are collected over the Internet. Internet surveys can substantially reduce the cost of reaching potential respondents and offer some of the advantages of in-person interviews by allowing the computer to show the respondent pictures or lists of response choices in the course of asking the respondent questions. The key limitation is the lack of control on the representativeness of the sample and self-selection bias.

- *In-person interviews* are the most common form in India. Though they are much more expensive than mail or telephone surveys, they enable collection of much more complex and detailed information. Furthermore, they not only allow the researcher more control over the sample population, but also, if well constructed, less sampling errors.

Some surveys combine various methods. For instance, a survey worker may use the telephone to "screen" or locate eligible respondents and then make appointments for an in-person interview.

In order to be effective, surveys need to be:

- ***Clearly defined*** *— before beginning, you need to be able to state the goals and objectives*

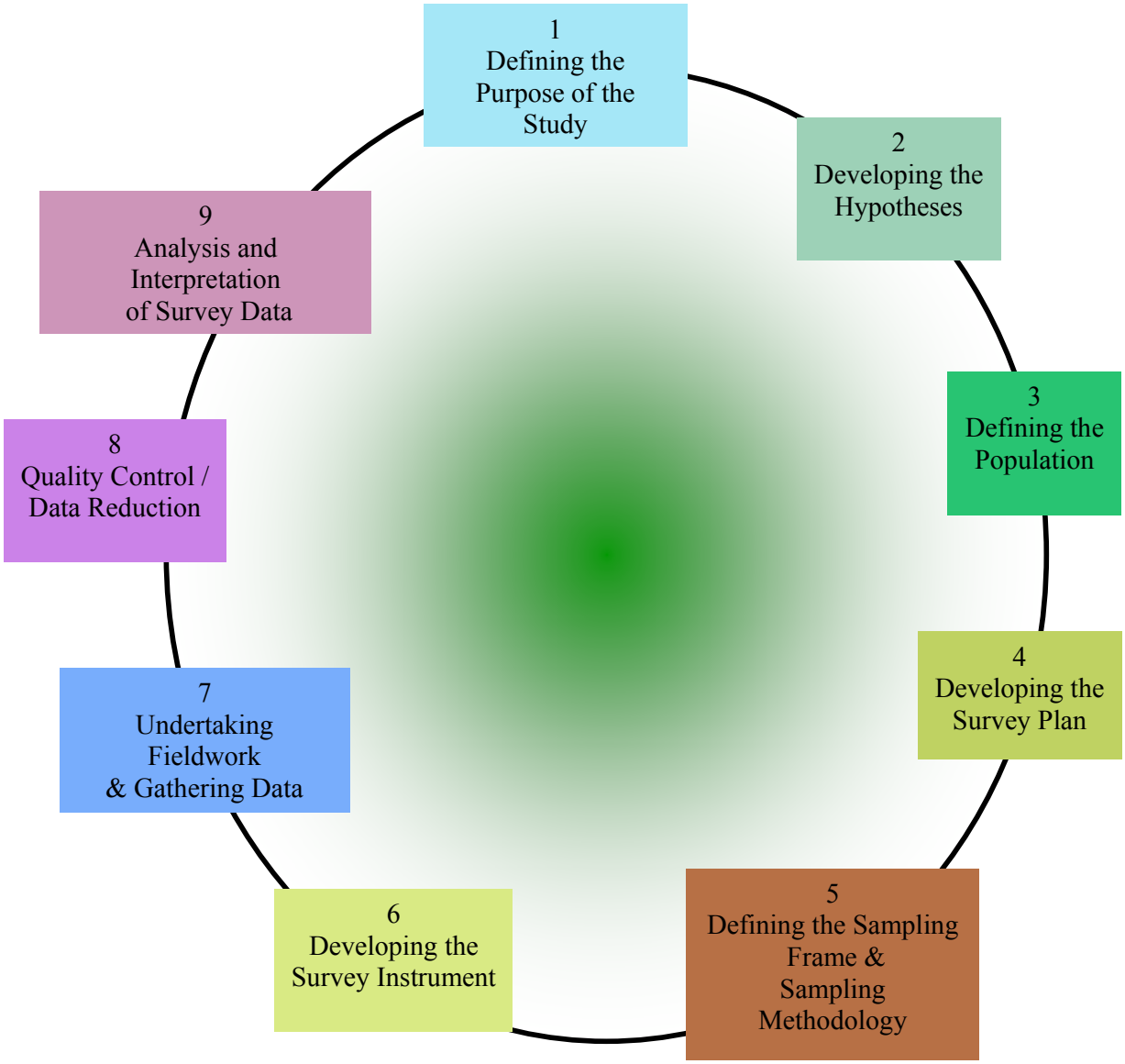
- ***Easily completed*** *— the respondents must be able to easily understand and follow your questions*

- ***Smoothly processed*** *— before you can begin analysis, the data must be clean and valid*

- ***Thoroughly analysed*** *— to get useful, reliable results, you need to be able to thoroughly analyse your data*

- ***Timely*** *— the time between planning and deployment must be short enough to make a difference*

Steps involved in a Survey
The following steps are involved in a survey exercise.



4.2 Beginning the Survey Process

Step 1: Defining the Purpose of the Survey
The first step in producing a survey is to define the purpose or objective of the survey. A clear statement of purpose is necessary not only as a justification of the project, but also as a guideline to determine whether future actions in the project are in support of the original purpose. Knowledge of the exact nature of the problem (objective) would determine exactly what kind of data to collect and what to do with it. It is imperative, at the outset, to ensure that:

a. the problem is well stated
b. the stated problem is the real problem
c. the surveyor understands exactly what the problem is

The survey should be designed to answer only the stated problem. Adding additional *interesting* objectives will lengthen and complicate the survey while clouding the real issue.

Step 2: Developing the Hypotheses
Once the problem has been clearly stated, the next step is to form one or more hypotheses. The hypothesis is actually an educated guess about the answer to the problem. It ought to be based on prior experience related to the problem, or based on any knowledge one may have of previous research done on the topic. Without such a framework in which to make an educated guess, there is no basis for making a guess at all. If there is no clear basis for formulating a hypothesis, one should instead develop one or more objectives or questions to frame the scope of the questionnaire.

It is important, at this stage, to point out that any hypothesis must be supported by credible evidence. Using anecdotal evidence to frame hypotheses may severely compromise on the nature of the survey.

An example in this context may better illustrate the above. Anecdotal evidence may suggest that a particular scheme in X district has not yielded economic benefits for the poor due to the presence of intermediaries. Formulating a hypothesis on the basis of the anecdotal evidence available would lead one to naturally construct a questionnaire aimed at assessing the programme in this context while overlooking other aspects. In other words, establishing the hypothesis may blind you to collecting data on other possible causes of the problem. This is the reason why hypotheses must be backed by a solid base in theory or previously gathered evidence that suggests the hypothesis is, in fact, probable.

Hypotheses must also be carefully written. They should not contain moral judgments or biased statements such as "All politicians are good leaders." There are many ideas on what constitutes a good leader and one person's idea may not be the same as that of others. One needs to avoid words like should, best, good, bad, and ought.

Hypotheses should be as specific as possible. Ambiguous words such as most and some should be avoided. A survey can more easily be designed to test whether "more than 75% approve" than whether "most approve."

A well-formulated hypothesis, objective, or research question translates the purpose into a statement that can be investigated scientifically. Without well formulated hypotheses, producing a valid survey becomes a very difficult task indeed.

Step 3: Defining the Population
It is important at this stage to identify the population or the target group that one is interested in. This is likely to emerge from the purpose of the survey and the hypotheses formulated.

Not only is it important to identify the population but one should endeavour to define the target segment as well as possible. For this purpose, one could choose many different criteria such as:

- geographical (ex: districts, hills, plains, agroclimatic zones, etc.)
- demographic (ex: urban/rural, age, sex, etc.)
- socio-economic (ex: APL/BPL, monthly income/expenditure, type of housing, castes/class etc.)
- other (such as attitudinal & behavioural characterictics, etc.)

A well-defined target segment lends well to the subsequent tasks of defining the sampling frame and adopting a robust sampling methodology to reach the target segment so defined.

Step 4: Developing the Survey Plan
The next step in the survey process is construction the survey plan. The purpose of the survey plan is to ensure that the survey results will provide sufficient data to provide an answer (solution) to the problem being investigated. The survey plan is comprised of:

- survey methodology
- data collection plan
- data reduction and reformatting plan
- analysis plan

None of these plans stand on their own – they are interlinked and interlocked with one another. How the data will be analysed had implications on the data collection plan. The type of data reduction planned will affect not only the types of analyses, but also the amount and types of data that will be needed to be collected. Given the fact that these plans are closely interrelated, they should be developed concurrently.

Survey Methodology
This involves determining the broad nature of the study. Should the study be a one time cross-sectional study or should it be done at regular time intervals? Such a decision will have implications on the eventual sample design and data collection plan.

The Data Collection Plan
The purpose of the data collection plan is to ensure that proper data are collected in the right amounts. The appropriateness of the data is determined by your hypothesis

and your data analysis plan. For example, if you plan to analyse your results by age group to test a hypothesis, then you must collect data from each age group.

As pointed out earlier, the use of sample data involves risk, and the amount of that risk is determined by the size of your sample. The amount of risk one is willing or able to accept should be stated in the analysis plan. Thus, one not only needs to collect data from some members of each group that one plans to analyse, but one also has to ensure that each group provides a response rate that is high enough to meet the minimum risk level.

The Data Reduction and Reformatting Plan
The purpose of the data reduction and reformatting plan is to identify upfront and to decrease as much as possible the amount of data handling. This plan is highly dependent on the other two plans. Proper coding of questions (both open-ended and close-ended) before the questionnaires are administered enable quick and error-free data reduction. Use of computers and statistical packages such as SPSS/SAS can be made in this regard. This function is ideally suited for outsourcing - there are numerous private agencies which can help in data reduction and reformatting. A strong potential for error and tedious corrective work lies in data reduction and reformatting. Proper care in developing this plan can save a great deal of time later and preclude error.

Analysis Plan
Finally, an analysis plan ensures that the information produced by the analysis will adequately address the originally stated hypotheses, objectives, or questions. It also ensures an analysis that is compatible with the data collected during the survey. The analysis plan determines which statistics one will use and how much risk one can take in stating your conclusions. Each of these decisions will affect the amount and type of data to be collected and it will be reduced. The most often committed error in statistical analysis is using a statistical technique with inappropriate data. A well designed analysis plan at the inception stage not only ensures robust and accurate statistical analysis but also helps save much time and effort at the data analysis stage.

One can simply approach the natural sequence of survey operations in reverse order. First determine what conclusions you are interested in; then decide what statistics and results will be needed to draw these conclusions. From this, the type of questions needed and the nature of the sample can be determined. A conscientious survey plan will help you produce a well designed survey. The *proper* data will be processed correctly and efficiently to produce the information required, and hopefully provide a solution to, the original problem.

4.3 Developing a Robust Sampling Methodology

Step 5: Determining the Sampling Frame and Sampling Methodology
When undertaking any survey, it is essential to obtain data from people that are as representative as possible of the group that one is interested in. Even with the perfect questionnaire (if such a thing exists), the survey data will only be regarded as useful if it is considered that respondents are typical of the population as a whole. For this reason, an awareness of the principles of sampling is essential to the implementation of most methods of research, both quantitative and qualitative.

Some key definitions:

Population: The group of people, items or units under investigation

Sampling Frame: The list of people from which the sample is taken. It should be comprehensive, complete and up-to-date. *Examples of sampling frame*: Electoral Register; Postcode Address File; telephone book

Sampling: it is the process of selecting a proper subset of elements from the full population so that the subset can be used to make inference to the population as a whole.

The Law of Statistical Regularity: A reasonably large sample selected at random from a large population will be, on average, representative of the characteristics of that population.

The Law of the Inertia of Large Numbers: Large groups of data show a higher degree of stability than smaller ones; there is a tendency for variations in the data to be cancelled out by each other.

Normal Distributions: They are a family of distributions that have the same general shape. They are symmetric with scores more concentrated in the middle than in the tails and, therefore, assume the shape of a bell-shaped curve.

Sampling and Sampling Errors

The crucial factor in making a survey successful is reducing "error." Survey error is the term used to describe any reasons that interfere in collecting perfect results. There are two types of survey error: a) non-sampling error and b) sampling error. Both can be controlled.

Non-sampling error results from poor questionnaire construction, low response rates, non coverage (missing a key part of the market), and processing weaknesses. The other type of error is sampling error. Sampling is the process of deciding what portion(s) of your universe will be surveyed, including who and how many. The goal of sampling techniques is to reduce (or eliminate) sampling error. In the ideal world, you wouldn't need sampling, and there would be no sampling error. One would (and could) survey all units of your population (called a census However, if your pool of respondents is large, hard-to-reach, or otherwise problematic, the only approach is to use a sampling technique. The following chart compares the benefits of sampling and census techniques.

Benefits of Sampling and Census

Sampling	Census
Lower Cost	Greater acceptance of results
Faster	Data for entire small populations may be obtained
More in-depth analysis possible	No random sampling error

Practicality	Does not require the use of complex theory to select a representative sample or in analysing results
Greater confidentiality	Data may be reported on every segment in the population
Greater accuracy	Subtle differences become apparent

There are two basic types of sampling errors – systemic and random. Systemic errors occur when the sample selected reflects a bias, in other words, does not reflect the range of findings for the universe. Systemic error can be greatly reduced by carefully estimating the universe – what the key segments are their relative sizes. Random error is the other sampling error – and the most common. It relates directly to the size of the sample – and is basically a mathematical predictor of precision. A general rule of thumb: as sample size increases, random sampling error decreases. However, A carefully selected small sample can be more accurate than a less-carefully selected large sample.

Sampling Methodology

The basic steps in selecting a sample are as given below:

- Define the universe. Who do you want to get information from? Decide the units (say BPL households), the elements (adult members), the extent (benefited from a scheme), and time (in the last one year). These factors put a limit on the survey.

- Develop a "sampling frame." Who are the people that make up the group(s) you want to survey? In the above example, the list of all BPL households will serve as the sampling frame for sampling of households.

- Specify the sampling unit and element. What specific segment(s) will get you the information you need? Say adult members within BPL households, who may or may not have benefited from the scheme

- Specify sampling method. What selection criteria will you use: probability vs. non-probability?

Probability sampling means that every segment of the population will most likely be included in a typical sample. Non-probability sampling is selection based on the researcher's judgment or convenience.

Types of Probability Sampling

- A simple random sample is one in which each member (person) in the total population has an equal chance of being picked for the sample. In addition, the selection of one member should in no way influence the selection of another. Simple random sampling should be used with a homogeneous population, that is, one composed of members who all possess the same attribute you are interested in measuring.
- Systematic sampling involves collection of a sample of survey participants systematically where every Kth member is sampled in the population where K is equal to the population size divided by the required sample size.

- Random Route Sampling Address is selected at random from sampling frame (usually electoral register) as a starting point. The interviewer is then given instructions to identify further addresses by taking alternate left- and right-hand turns at road junctions and calling at every nth address.

- A stratified random sample is defined as a combination of independent samples selected in proper proportions from homogeneous groups or strata within a heterogeneous population. In other words, all people in sampling frame are divided into "strata" (groups or categories). Within each stratum, a simple random sample or systematic sample is selected.

- Multi-stage cluster Sampling involves drawing several different samples. Initially large areas are selected and then progressively smaller areas within larger area are sampled. Eventually, this ends up with a sample of households.

Types of Non-Probability Sampling
It isn't always possible to undertake a probability method of sampling, such as in random sampling. In such situations, a non-probability sampling technique may be adopted.

- Purposive sampling is one in which respondents are selected by the researcher subjectively. The researcher attempts to obtain sample that appears to him/her to be representative of the population and will usually try to ensure that a range from one extreme to the other is included.

- Quota sampling is often used to find cases with particular characteristics. Interviewers are given quota of particular types of people to interview and the quotas are organised so that final sample should be representative of population.

- A convenience sample is one that comprises subjects who are simply available in a convenient way to the researcher. This could be at a crossroads, shopping mall or street corner.

- In Snowball Sampling potential respondents are contacted and then they provide information on other potential respondents with the same characteristics who are then contacted.

- Self-selection is perhaps self-explanatory. Respondents themselves decide that they would like to take part in your survey.

Determining Sample Size
There are four key considerations that determine sample sizes of a survey.

Population Size: In other words, how many people are there in the group that the sample represents? This may be the number of people in the state/district/town you are studying, or the number of BPL households or the electoral population, as the case may be. Often the exact population size may not be known. This is not a problem. The mathematics of probability proves the size of the population is irrelevant, unless the size of the sample exceeds a few percent of the total population that one is examining.

Sampling Risk: The less risk you are willing to take, the larger the sample must be. Risk, as it relates to sample size determination, is specified by two interrelated factors:

- the confidence level
- the precision (or reliability) range.

The confidence level tells you how sure you can be. It is expressed as a percentage and represents how often the true percentage of the population who would pick an answer lies within the confidence interval. The precision range (standard error) reflects the deviation of the sample estimate from the actual population value. To minimise risk, one should have a high confidence (say 95%) that the true value you seek (the actual value in the population) lies somewhere within a small interval (say + or – 5%) around your sample value (your precision).

Analysis Plan: Another factor bearing on sample size is also obtained from your analysis plan. If there are many sub-groups covered within the population, the sample size requirements may be larger than from a homogeneous population. Similarly, if the study mandates accurate reporting at a sub-group/strata level, adequate sample sizes would need to be provided at each stratum/subgroup level.

Time and Cost: Inadequate time or high costs often curtail sample sizes of a survey. In such circumstances, the confidence level of reporting and standard error of estimation are compromised.

Sample sizes can be estimated using the following formula:

$$n = \frac{NZ^2 * .25}{[d^2 * (N-1)] + [Z^2 * .25]}$$

- **n** is the sample size
- **N** is total population size (known or estimated)
- **d** is the desired precision/margin of error
- **Z** is the value of corresponding the desired confidence level obtained from a normal distribution table (usually 95%)

The above assumes the worst case scenario where the sample proportion (p) has been assumed to be 0.5. Hence [p * (1-p) = 0.25]. This yields the maximum sample size required to report for a variable of interest at a predetermined confidence level allowing for a certain margin of error.

4.4 Developing an Effective Survey Instrument/Questionnaire

Step 6: Questionnaire Design

Questionnaires play a central role in the data collection process. The questionnaire is the means for collecting your survey data. A well-designed questionnaire efficiently collects the required data with a minimum number of errors. It facilitates the coding and capture of data and it leads to an overall reduction in the cost and time associated with data collection and processing.

A poorly constructed questionnaire can invalidate a robust survey design as it gives rise to non-sampling error. The key to minimising the disadvantages of the survey questionnaire lies in the construction of the questionnaire itself. Since the questions are the means by which you are going to collect your data, they should be consistent with your survey plan. The biggest challenge in developing a questionnaire is to translate the objectives of the data collection process into a well-conceptualised and methodologically sound study. Properly constructed questions and well-followed survey procedures will allow you to obtain the data needed to check your hypothesis and, at the same time, minimise the chance that one of the many types of bias will invalidate your survey results.

Questionnaire design as a task must not be seen in isolation from other aspects of survey design. A well constructed instrument must take into account:

- Survey objectives: The objectives of the survey lead to identification of areas on which information would be required which, subsequently, translate into questions to be asked to respondents.

- The type of survey: Different types of surveys may have different implications on the questionnaire. An opinion poll would necessitate more scaled responses rather than a programme assessment survey. Similarly, a beneficiary identification survey may be much shorter than a survey to assess municipal services.

- The target segment: The target segment for the survey may consist of heterogeneous groups which are differentiated in socio-economic and cultural characteristics. The questionnaire structure (and wording, in some cases) is often customised for different homogeneous segments within the population.

- The sampling methodology: The questionnaire should capture the basic strata/group codes on which the sampling methodology has been based. This enables correct weighting of the data at a later stage.

- The analysis plan: The analysis plan envisages analysis of the data along particular criteria. The questionnaire should be, therefore, constructed in a manner that conforms to the analysis plan.

- Cost and time constraints: Finally, cost and time constraints may influence the questionnaire in terms of the both the depth and the width of information to be collected.

The Art of Questionnaire Design

Firstly, the questionnaire should have a uniform structure. The questionnaire should be organised in such a way that:

- Demographic questions come at the beginning/end
- opening questions arouse interest.
- easier questions are asked first.
- general questions precede specific ones.
- similar questions are grouped together.
- personal or emotional questions should be kept to the end.

The introduction of the questionnaire is very important because it outlines the pertinent information about the survey being conducted. The opening questions of any survey questionnaire should establish the respondents' confidence in their ability to answer the remaining questions. The introduction should:

- provide the title or subject of the survey;
- identify the sponsor;
- explain the purpose of the survey;
- request the respondent's co-operation; and
- inform the respondent about confidentiality issues, the status of the survey (voluntary or mandatory) and any existing data-sharing agreements with other organisations.

Types of Questions

Before investigating the art of question writing, it will be useful to examine the various types of questions. Broadly, four types of questions are used in surveying.

- The *background question* is used to obtain demographic characteristics of the group being studied, such as age, sex, grade, level of assignment, and so forth. This information is used when to categorise results by various subdivisions such as age, income, caste, religion, etc. Therefore, these questions need to be consistent with the analysis plan.

- The second type is the *open-end question*. This type requires respondents to answer the question in their own words. The open question allows the respondent to interpret the question and answer it anyway he or she chooses and can be used to gather opinions or to measure the intensity of feelings.

- The third and most common type of question is the *closed-end question*. It is used to determine feelings or opinions on certain issues by allowing the respondent to choose an answer from the list provided. For the respondent, a closed question is easier and faster to answer and for the researcher, closed questions are easier and less expensive to code and analyse. Also, closed questions provide consistency, an element that is not necessarily associated with an open question.

- The *intensity question*, a special form of the multiple-choice question, is used to measure the intensity of the respondent's feelings on a subject. These

questions provide answers that cover a range of feelings and are usually in the form of scaled responses.

Each type of question has its own strengths and weaknesses. A good questionnaire makes judicious use of different types of questions given the nature and type of information requirements.

When writing questions, the following should be kept in mind:

- The language should be simple. There is need to avoid jargon and complex terminology.

  For example: “Do you know about the impending plans for convergence of different poverty alleviation schemes?”

  Better wording: “Do you know that there are plans being made to bring together different poverty alleviation schemes?”

- The frame of reference for a question should be clear.

  For example: "What is your income?"

  Does the word "your" refer to the respondent's personal income, family income or household income? Does the word "income" refer to salary and wages only, or does it include income from other sources? Because there is no specific time period mentioned, does this question refer to last week's income, last month's or last year's income?

- Some questions may need to state the type of response needed.

  For example: Respondents are shown a bottle of orange drink and are asked, "How much orange juice do you think this bottle contains?”

  The answers obtained could be much varied such as:

  “One orange and a little water and sugar”
  “25% orange and 75% carbonated water”
  “Juice of one-half dozen oranges”
  “Three ounces of orange juice”
  “Full strength”
  “A quarter cup of orange juice”
  “None”

  Better wording: "This bottle holds 250 millilitres (mL) of orange drink. How many millilitres of this drink would you say are orange juice?"

- The questionnaire should not contain double-barreled questions (a question that asks two questions simultaneously rather than one)

For example:

"Are you happy with your elected representative and will vote for him in the next elections?"

"Do you plan to leave your car at home and use public transport to work during the coming year?"

In some instances, the response for each half of the question is the same. However, many other responses could include two very separate answers which would make interpreting this question difficult.

- Loaded questions corrupt the survey findings. Hence, questions should be neutral and unbiased.

  The effect of loaded questions can be gauged from the following hypothetical example.

  "In your opinion, should key municipal services be outsourced to more efficient private agencies?"
  Results:
  83% In favour of outsourcing
  14% Opposed to Sunday shopping
  3% No opinion

  "In your opinion, should key municipal services be outsourced to private agencies who charge higher user fees?"

  Results:
  62% opposed to outsourcing
  32% In favour of outsourcing
  6% No opinion

  In both the above cases, respondents react not to `outsouring' per se but to the perceived impact of outsourcing on them.

- Questions should not seek information that may be inaccessible from the respondents.

  An item may use familiar terms but require information most respondents would not know. For instance "Is your family income above, about equal to, or below he official poverty rate?" is flawed because people are not apt to know what the official poverty rate is, making the item unacceptable for a determination of fact (although possibly acceptable for a determination of self-perception).

In addition to the above, the questionnaire should allow for the following:

- The number of questions should be kept to an optimal limit.
- The questions should be short and easy to follow.

- Questions should allow for all possible answers.
- It should include a few questions that serve as checks on the accuracy and consistency of the answers as a whole.

Pre-testing/Piloting the Instrument

This is a fundamental step in developing a questionnaire. The purpose of the pretest is to assess every key detail involved in the questionnaire design namely, the interview time, the ease of comprehension of instructions, interpretation of questions, the type of answers a stimuli evokes, interviewer/respondent fatigue, etc.

Testing helps discover poor wording or ordering of questions; identify errors in the questionnaire layout and instructions; determine problems caused by the respondent's inability or unwillingness to answer the questions; suggest additional response categories that can be pre-coded on the questionnaire; and provide a preliminary indication of the length of the interview and any refusal problems. Testing can include the entire questionnaire or only a particular portion of it. A questionnaire will at some point in time have to be fully tested.

The following is a list of some key points to think about when designing the questionnaire:

- Is the introduction informative? Does it stimulate respondent interest?
- Are the words simple, direct and familiar to all respondents?
- Do the questions read well? Did the overall questionnaire flow?
- Are the questions clear and as specific as possible?
- Does the questionnaire begin with easy and interesting questions?
- Does the question specify a time reference?
- Are any of the questions double-barreled?
- Are any questions leading or loaded?
- Should the questions be open- or close-ended? If the questions are close-ended are the response categories mutually exclusive and exhaustive?
- Are the questions applicable to all respondents?

4.5 Data Collection and Analysis

Step 7: Undertaking Fieldwork and Gathering Data
This is the first operation part of the survey process. A well designed sampling methodology must be complemented by good standards in the actual gathering of data through professionally trained investigators.

For conducting citizen surveys, government departments are advised to use the services of professional research agencies that have strong field operations and professional investigators. However, they should ensure that the process of data collection by the agencies concerned subscribes to the following:

- Operational planning: This is meant to serve as a roadmap for the actual survey. This incorporates resource planning in order to align manpower to the survey design and time constraints. Use of activity charts can be a useful method of planning fieldwork operations.

- Training of investigators: Given that most surveys in India are personal face-to-face interviews, this aspect assumes great importance. It is important for investigators, who undertake the work of interviewing respondents, to clearly understand the purpose of the survey and the target respondent. They should be aware of the reason for each question in the instrument. Investigators should also know the micro-level sampling methodology on the basis of which they would have to select the area, the household and the respondent within the household. In this regard, use of investigators who are familiar with such surveys may be an advantage.

- Monitoring and supervision: Mechanisms should be in place to adequately monitor and supervise the fieldwork operations. This has a bearing on both the time and quality of the survey. Proper monitoring of the field teams can help to regulate and control the progress of fieldwork. Interview accompaniments and backchecks serve as an important means for ensuring good quality of interviewing.

Step 8: Quality Control / Data Reduction

Data preparation and management
The goal of the data preparation and management stage is to get the data ready for analysis. When examining a new data set, data verification and cleaning ensures that the analytical results are accurate. For example, if there is gender data in which "1" is for male and "2" is for female, the data shouldn't have "3" as a response. Using data collection software during this step can help streamline the process.

Setting up the "codebook"
During the data preparation and management step, the first step is to set up a "codebook" information, which is any variable definition information. This includes variable names, variable formats and descriptive variable labels (data such as gender or income level) and value labels (numbers assigned to data, such as "1" for male, "2" for female).

Setting up multiple-item indices and scales
Multiple-item indices and scales, which combine multiple indices into a single, multiple-item index can also be set up. This provides a more reliable measurement of interest than a single question can. This will enable better cross-tabulation and multiple-item analysis.

This stage also involves:
(a) Transformation of data: This helps to get your data in a structure and form needed for analysis.
(b) Filling in missing data: Replacing missing data values with estimates ensures better summary statistics.

Step 9: Analysis and Interpretation of Survey Data

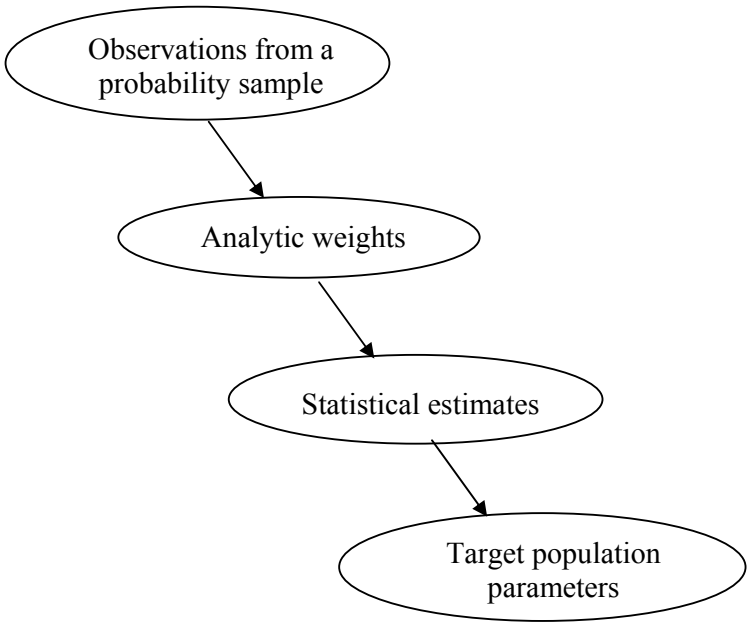
Weighting of data
Before analysing and interpreting the data, it may be required to `weight' the data. Weighting refers to the construction of a weight variable. The principal purpose of weighting is to obtain as accurate parameter estimates as possible with the chosen sampling and estimation procedures.

The simplest type of surveys may be "self-weighted" in the sense that each unit (household/group) in the survey "represents" the same number of unit in the population. Some surveys are close to being self-weighted because they do not deliberately oversample any particular sub-group in the population but instead draw the sample in a way that each unit (individual/households) in the population has the same probability of being selected in the sample. Yet variation in response rates across different types of households/groups usually implies that weights must be calculated to correct for such variation. Indeed, most surveys are not self-weighted because they draw disproportionately large samples for some parts of the population that are of particular interest. In this case, weights must be used when presenting all descriptive statistics in order to calculate unbiased estimates of statistics of interest.

Accurate weights must incorporate three components. The first is the "base weights". These account for variation in the probabilities of being selected across different groups of households as stipulated by the survey's initial sample design. The second is adjustments for variation in non-response rates. For example, in many developing countries wealthier households are less likely to agree to be interviewed than are middle income and lower income households. The base weights need to be "inflated" by the inverse of the response rate for all groups of households. Finally, in some cases there may be "post-stratification adjustments". The basic idea here is that some other data source, such as a census, may provide very precise estimates of the distribution of the population by age, sex, and ethnic group. If the survey estimates of these distributions do not match those given my the other, more accurate data source, the survey data should be re-weighted so that, with the new weights, the survey reproduces the distributions from the other data source.

The final analytic weights attached to each analytic file produced from a survey may contain the following factors:

- The design-based weight computed as the reciprocal of the overall probability of selection;
- A non-response adjustment factor;
- A post stratification adjustment factor;
- A weight-trimming factor



Application of Analytic Weights and Statistical Estimation

Types of Data

Nominal	Ordinal	Interval	Ratio
People or objects with the same scale value are the same on some attribute. The values of the scale have no 'numeric' meaning in the way that you usually think about numbers.	People or objects with a higher scale value have more of some attribute. The intervals between adjacent scale values are indeterminate. Scale assignment is by the property of "greater than," "equal to," or "less than."	Intervals between adjacent scale values are equal with respect to the attribute being measured. E.g., the difference between 8 and 9 is the same as the difference between 76 and 77.	There is a rationale zero point for the scale. Ratios are equivalent, e.g., the ratio of 2 to 1 is the same as the ratio of 8 to 4.
Ex: Gender, marital status	Ex: any ranking	Ex: personality measurements, Degrees C	Ex: Length or distance in centimeters, inches, etc.

Data Analysis

Data analysis enables the extraction of useful information from the collected data which leads to informed decision-making. Every piece of the acquired data has intrinsic value. The key is extracting this value. One is able to better understand the target segment, whether it's customers, employees or citizens, by analysing the most intimate details. Different statistical procedures are appropriate for variables depending on what knowledge is required and the level of measurement of the variable.

Broadly, analysis of data could be categorised into two types. Descriptive data analysis helps in organising and summarising data in a meaningful way. Description is an essential step before any further statistical analyses. The goals of descriptive data analysis are to (a) summarise data and (b) get an accurate description of the variables of interest. Inferential data analysis allows the researcher to make decisions or inferences by identifying and interpreting patterns in data. Inferential statistics deal with drawing conclusions and, in some cases, making predictions about the properties of a population based on information obtained from a sample. While descriptive statistics provide information about the central tendency, dispersion or skew, inferential statistics allow making broader statements about the relationships between data.

The primary purpose of the vast majority of sample surveys, both in developed and in developing countries, is descriptive, although there is increasing interest in making inferences about the relationships among the variables investigated. A common error, often committed by researchers, is use of inferential statistics without a robust a priori model survey design. The use of descriptive analysis is, often, underestimated - simple basic information on the variables in the form of descriptive statistics may be more valuable than complex relationships revealed through half-baked inferential statistics. This paper, while elaborating on the former, also provides a list of techniques that could be used to make inferences from the datasets.

Descriptive Data Analysis

Most citizen survey data can be used in a wide variety of ways to shed light on the phenomena that are the main focus of the survey. In almost all cases the starting point is basic descriptive statistics such as tables of the means and frequencies of the main variables of interest. The first step in any data analysis is to generate a data set that has all the variables of interest in it. Some variables may come directly from the "raw data" without any need for modification, such as the sex of the head of household, while other variables will have to be generated by transforming the raw data, such as calculating net income from farming activities by using detailed information on crop sales and purchases of agricultural inputs. Once the variables to analyse have been created and put into a single data set, the first task is to generate basic descriptive statistics that let the variables "speak for themselves".

Any variable can be classified into one of two types, discrete variables and continuous variables. Discrete variables take only a small number of values. For example, the type of flooring in a household's dwelling can be dirt, wood, concrete, or some other type of material. Continuous variables can take an infinite, or almost infinite, number of values, such as the amount of land that a rural household may own. Methods for presenting data are different for these two types of variables.

Descriptive Data Analysis

Descriptive analysis can done in many ways viz. the number and type of variables used to present the data.

One Variable

1. Discrete variables: Discrete variables are the simplest to display. As long as the number of values that the variable can take is reasonably small, say ten or less. One simply can display that variable in terms of the frequency that if takes each of those variables through tables, pie-charts or graphs.

2. Continuous Variables: Continuous variables can be displayed in many different ways. First, one can divide the range of any continuous variable into a discrete number of intervals and display the information in any of the ways that can be used for discrete variables. For example, the income of households, which can take an infinite number of values, could be divided into a small number of categories, such as less than Rs 2,000 per month or less, between Rs. 2,000 and Rs. 5,000 per month, etc., with the final category being more than Rs. 50,000 per month. There are different ways in which they could be displayed such as summary statistics like mean and standard deviation, minimum and maximum values. The distribution could also be highlighted through histograms.

Two Variables.

Comparing two or more variables often offers much more insight into the underlying topic of interest than examining a single variable in isolation. Yet at the same time the possibilities for displaying the data increase by an order of magnitude.

1. Two discrete variables: The simplest case for displaying the relationship between two variables is that where both variables are discrete and each takes a small number of discrete values. In a simple two-way tabulation, the values of one variable can serve as the columns while the values of the other variable can serve as the rows. There are several ways to display information on the relationship between two discrete variables such as row percentages or column percentages or table percentages.

2. One continuous and one discrete variable: Here the most common way to display the data is in terms of the mean of the continuous variable conditional on each value of the discrete variable. Another option is to transform the continuous variable into a discrete variable be dividing its range into a small number of categories. For example, it is sometimes convenient to divide households into the poorest 20%, the next poorest 20%, and so on, based on household income or expenditures. After this is done, one can use the same methods for displaying data for two discrete variables, as described above.

3. Two continuous variables: Statisticians often provide summary information on two variables in terms of their covariance or their correlation coefficient. However, such statistics are often unfamiliar to a general audience. An alternative is to graphically display the data in a scatterplot that has a dot for each observation. This could show, for example, the extent to which household income is correlated over two periods of time. Sometimes information on the joint distribution of two or more variables may be clearest if one or both variables is transformed into a discrete variable, such as

transforming income into five "quintile" categories. This is often easier for a wide audience to understand.

Three or more variables
For three variables, the most straightforward approach is to designate one variable as the "conditioning" variable. This variable with either have a small number of discrete values or, if continuous, it will have to be "discretised" by calculating its distribution over a small number of intervals over its entire range. After this is done, separate tables or graphs can be constructed for each value of this conditioning variable. For example, suppose one is interested in showing the relationship between three variables, the education of the head of household, the income level of the household, and the incidence of child malnutrition. This could be done by generating a separate table or graph of the relationship between income and child nutrition for each education level. This may show, for example, that the correlation between income and child nutrition is weaker for households with more educated heads.

Inferential Data Analysis
The two major types of inferential statistics are parametric statistics and non-parametric
statistics.

Parametric Tests

- assume that the variable measured is normally distributed in the population
- the data must represent an interval or ratio scale of measurement
- the selection of participants is independent
- the variances of the population comparison groups are equal

Non-Parametric tests are less powerful means of data analysis and are used when the data represent a nominal or ordinal scale, when a parametric assumption has been greatly violated, or when the nature of the distribution is not known.

Some common parametric tests

t-test: This is used to determine whether two means are significantly different at a selected probability level. The strategy of the t-test is to compare the actual mean difference observed to the difference expected by chance. If the t value is equal to or greater than the table value, then the null hypothesis is rejected because the difference is greater than would be expected due to chance. This test can be done for both independent samples (randomly formed) and for non-independent samples (non-randomly formed).

ANOVA (Analysis of Variance): This is used to determine whether two or more means are significantly different at a selected probability level and thus avoids the need to compute duplicate t-tests to compare groups. The strategy of ANOVA is that total variation, or variance, can be divided into two sources: a) treatment variance ("between groups") and error variance ("within groups"). If the treatment variance is sufficiently larger than the error variance, a significant F ratio results, that is, the null hypothesis is rejected.

Pearson Correlation Coefficient: To express a relationship between two variables one usually computes the Pearson correlation coefficient. It measures the linear relationship between two interval/ratio level variables. This is defined as the ratio of the joint variation of the independent and dependent variables relative to the variation in both variables considered separately. Pearson's r is always between -1 and +1, where -1 means a perfect negative, +1 a perfect positive relationship and 0 means the perfect absence of a relationship.

Some common non-parametric tests

Chi Square: The Chi Square (X2) test is undoubtedly the most important and most used member of the nonparametric family of statistical tests. Chi Square is employed to test the difference between an actual sample and another hypothetical or previously established distribution such as that which may be expected due to chance or probability. This a nonparametric test of significance appropriate for nominal or ordinal data that can be converted to frequencies. It compares the proportions actually observed (O) to the proportions expected (E) to see if they are significantly different. The chi square value increases as the difference between observed and expected frequencies increases.

Mann-Whitney Test: This test is a non-parametric independent samples test for the difference in central tendencies of two populations. The Mann-Whitney test is performed by combining the two data sets we want to compare and calculating their mean ranks to see if the sample data is in the correct direction. Unlike the t-test, it does not require us to assume that the dependent variable is normally distributed or even measured at interval level, although it has less statistical power than a t-test.

Kolmogorov-Smirnov Test (KS-test): This is a goodness-of-fit test for any statistical distribution. The test relies on the fact that the value of the sample cumulative density function is normally distributed. To apply the Kolmogorov-Smirnov test, the cumulative frequency (normalised by the sample size) of the observations is calculated and compared to the the cumulative frequency for a true distribution. The greatest discrepancy between the observed and expected cumulative frequencies is called the D-statistic which is compared against the critical D-statistic for that sample size.

Spearman Correlation Coefficient: It is designed to measure the degree of relation for two ordinal variables. It can be used when the two variables are ranks or when either one or both variables are changed into ranks. Once the variables are converted into ranks, the same procedure as used for Pearson correlation is followed.

In addition to these tests, inferential statistics also includes many multivariate techniques. Multivariate techniques have the advantage of enabling the simultaneous analysis of two or more variables. Some key multivariate techniques and their purpose are given in the table below:

	Multivariate technique	Purpose of technique
1.	Principal component analysis	Dimension reduction by forming new variables (the principal components) as linear combinations of the variables in the multivariate set
2.	Cluster	analysis Identification of natural groupings amongst cases or variables
3.	Factor analysis	Modelling the correlation structure among variables in the multivariate response set by relating them to a set of common factors
4.	Multivariate analysis of variance (MANOVA)	Extending the univariate analysis of variance to the simultaneous study of several variates. The aim is to partition the total sum of squares and cross-products matrix amongst a set of variates according to the experimental design structure
5.	Discriminant analysis	Determining a function that enables two or more groups of individuals to be separated on the basis of multiple responses on all individuals in the groups
6.	Canonical correlation analysis	Studying the relationship between two groups. It involves forming pairs of linear combinations of the variables in the multivariate set so that each pair in turn, produces the highest correlation between individuals in the two groups
7.	Multidimensional scaling	Constructing a "map" showing a spatial relationship between a number of objects, starting from a table of distances between the objects

More information on one or more technique highlighted can be provided if it is solicited from interested departments.

All the tests highlighted can be easily done nowadays through the aid of special software packages such as SPSS, SAS, STATA, etc. However, before initiating such exercises, proper understanding of the technique, its limitations and the data requirements for the same must be ensured. The use of specialised agencies for data analysis is recommended especially in the case of citizen surveys that employ a complex survey design.

5 Case Study: A Citizen Survey on Corruption among Public Servants

This section will look at each step of the survey process through a hypothetical case study. This will enable a better understanding and appreciation of the survey process outlined in the earlier section.

Context:
The GoAP is acutely concerned about the level of corruption among public servants. It wishes to identify the departments that are perceived by citizens to be the most corrupt as well as assess the nature of such corruption. It wishes to undertake a citizens' survey for this purpose.

Given below is a step-by-step guide to how such a survey can be initiated, implemented and analysed for informed decision-making. This is simply a concise version aimed to supplement the theoretical approach with a practical example and hence has purely reference value.

Step 1: Defining the purpose of the survey

The key survey objectives could be:

Primary objectives

- to assess the incidence of corruption among public servants in their interface with citizens
- to assess the awareness among the public on specific initiatives taken by the government to combat corruption
- to identify the most corrupt departments as perceived by the citizens
- to assess the nature of corruption that citizens face in their interface with public servants of different departments viz.
    - type of corruption
    - type of officials
    - type of work concerned
    - amount paid
    - etc.

Secondary objectives

- to access the level of citizen participation in anti-corruption activities
- to access the degree and form of citizen reporting on corruption activities

Objectives should be spelt as clearly as possible. For example what does incidence of corruption mean. Is it in terms of bribes asked for by public servants from respondents themselves or in terms of respondents having seen bribes being paid by others or in terms of respondents knowing someone who has paid bribes?

Step 2: Developing the hypotheses

Each research objective lends itself to one or more hypotheses. Two key principles should be kept in mind while framing the hypothesis.

Remember: You never accept a hypothesis. You either reject or fail to reject it.

For example, suppose there is data to suggest that on an average incidence of corruption in India is 50%. The hypothesis, therefore, should be framed in a manner whereby the rejection of the hypotheses can only be done at a high confidence level (usually 95%). Otherwise, the hypothesis cannot be rejected.

Correct way
Ho: The incidence of corruption among public servants in Andhra Pradesh is higher than that in India taken as a whole i.e. greater than 50%.

Similarly, many more hypotheses can be framed on key variables of interest. Each such hypothesis, in conjunction with the research objectives, lends itself into specific information areas. Each information area would then feed into the survey instrument.

Step 3: Defining the population

Given the nature of the study, what is the target population? The research objectives must be kept in view while defining the population. In this case, it may be well advised to take all citizens (above the age of 15/18) as the universe. A tighter definition could exclude all citizens who have never interacted with any public servant.

Any decision in this regard must be made after careful deliberation given the fact that it may have implications for the outcome of the study.

Step 4: Developing the Survey Plan

Given the objectives of the study, a one-time cross-sectional design would be most appropriate. This would imply that the sample design proposed should represent all sections and the data collection plan should involve an extensive fieldwork programme over a defined time period. The findings from the study, however, will not enable pre and post analysis scenarios.

The analysis plan should ensure:

- data analysis by different demographic segments such as age, gender, urban/rural, districts, etc.
- feedback on different government departments
- composite indices of corruption based on citizen perceptions
- analysis of behavioural/attitudinal aspects

Step 5: Developing the Sampling Frame and Sampling Methodology

The sampling frame emanates from the target population definition. The electoral list that contains the names of all eligible voters could serve as a good sampling frame.

A multi-stage sampling methodology would be required for the survey. But prior to that the sample sizes need to be fixed at various levels as this would determine the confidence level of reporting and therefore the number of units to be sampled.

Suppose it is decided that the findings should be able to stand scrutiny at the district level. This implies that the accuracy of reporting should be within the desired

confidence level (say 95%) allowing for a specific margin of error (say 5%) at the district level. To fulfill this requirement, the sample sizes at the district level would turn out to be close to 400 (assuming 50% incidence of any variable of interest - the worst case scenario). Thus, the study would require a minimum of 9200 (23x400) as sample size.

This means that (a) reporting at the block/mandal level can be done only at a lower confidence level and (b) reporting at the state level would involve higher confidence and/or lesser margin of error.

Within each district, block/mandals to be covered can be stratified on the basis of any parameter thought to be fit, the assumption here being that similar blocks/mandals are homogeneous. In the absence of data to support stratification, simple random sampling of blocks/mandals can be done but this may increase sampling error.

Once blocks/mandals have been sampled, towns/villages within them would need to be selected. One could further stratify the towns.villages on the basis of population. Simple random sampling on the basis of probability proportional to sample size can be employed to sample towns/villages. The household selection within villages could be randomly done by following the right hand rule of field movement. (Different methodology may be adopted for villages). The random selection of an individual within the household for interview can be done using the random number grid.

Two key issues may be highlighted here. The number of blocks/ mandals/ towns/ villages to be sampled would depend on the sample sizes required and the minimum sample sizes required at each level as decided. Secondly, the sample sizes covered, if deviating from the proportions in the universe, will have to be weighted back.

Step 6: Developing the Survey Instrument

The information areas to be covered by the survey developed on the basis of research objectives serve as the primary basis of inputs for the survey questionnaire.

Given the nature of the study, different types of questions would need to be framed.

- Demographic questions: Questions regarding age, gender, locality, caste, etc. of respondent is important for getting two reasons. Firstly, they serve as parameters on which descriptive data analysis can be based. Secondly, it would highlight the nature of coverage and representation of the survey.
- Close-ended questions for behavioural data: These questions could be in the form of

  "Have you ever bribed a government official?"

  Yes: 1
  No: 2
  Don't Remember/Cant Say: 3

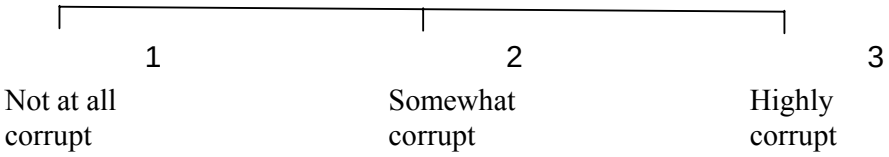
  "The last time you had to bribe any government official, how many rupees did you offer as bribe?"

Between 10 to 50 Rupees: 1
Between 51 to 100 Rupees: 2
Between 101 to 200 Rupees: 3
Between 200 to 500 Rupees: 4
Between 501 to 1000 Rupees:5
Between 1001 to 5000 Rupees: 6
Between 5001 to 10000 Rupees: 7
More than 10000 Rupees: 8

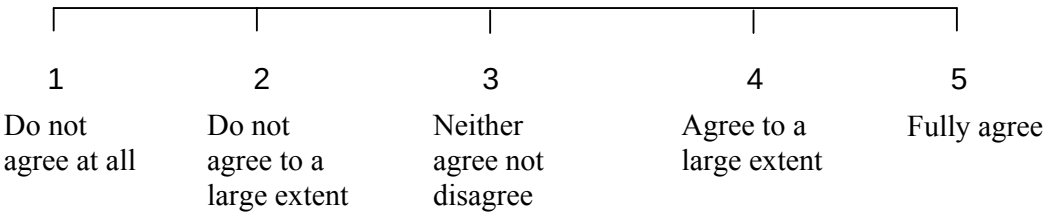
- Scaled questions: These questions would capture the perceptions of citizens on corruption. For example,

  "We would like to know your opinion on corruption among public servants. To what extent, do you think, are public servants corrupt? Please have a look at this card and answer."

  "Public servants are ……."



"To what extent do you agree with the following statements?"
"Please indicate the degree to which you agree or disagree with the following statements with the help of this card."



Sl. No.	Statement	Do not agree at all	Do not agree to a large extent	Neither agree not disagree	Agree to a large extent	Fully agree
1.	Corruption is widespread in Andhra Pradesh	1	2	3	4	5
2.	Corrupt government officials should be dismissed from service	1	2	3	4	5
3.	Citizens should play a greater role in anti-corruption strategies	1	2	3	4	5
4.	…….	1	2	3	4	5
5.	…….	1	2	3	4	5

- Open ended questions: Such questions would aim to get verbatim responses from respondents. For example,

  "What did you do when the government official asked you to pay a bribe?"

  or

  "Why didn't you lodge a complaint against the government official who asked you yo pay a bribe?"

The survey instrument should be piloted before it is finalised. This would help to make changes in the length of the questionnaire, specific questions, translation, etc.

Step 7: Undertaking the fieldwork

A total of 50 investigators and 10 supervisors can be used for this purpose. They can be made into 10 teams with each team consisting of 5 investigators and one supervisor.

On an average, each investigator can conduct 6 successful interviews. Thus 300 interviews can be conducted on an average every day. A total of 30-35 days would be required for the fieldwork to be completed. An additional 7-10 days must be kept for training of investigators and traveling. The travel planning must be done in the most efficient manner such that both time and cost can be kept to a minimum.

The supervisor must backcheck at least 25% of the questionnaires in order to vouch for the authenticity of the data. Additionally, the field officer must backcheck at least 5% of the questionnaires.

Step 8: Quality control/data reduction

The coding of data must be done in an efficient manner. In particular, the open-ended questions must be examined qualitatively before the responses are coded.

Missing data should be identified at this stage. Additionally, logic checks must be put in key places to ensure that the responses are robust.

Step 9: Analysis and Interpretation

The first step here is to weight the data. It is highly improbable that a survey of this nature will be self-weighted given the fact that stratified random sampling has been employed. Thus, in order to make the sample representative of the population, there would be need to weight the data.

Weighting data must be done keeping in mind the sampling and reporting considerations. In this case, the basic unit of reporting would be the district. Hence, the data must be weighted by urban/rural proportions at the stratified units level within the district (strata of mandals/blocks) in order to reflect the overall urban/rural proportions at the district level.

The subsequent analysis of the weighted data should be as per the analysis plan. Most of the analysis is likely to be descriptive in nature viz. incidence of corruption, perception of corruption by different target segments, forms of corruption, etc. Basic inferential statistics can also be employed such as correlation tests and variance analysis.