**JIMMA UNIVERSITY**

**COLLEGE OF NATURAL SCIENCE**

**DEPARTMENT OF STATISTICS**

Course Title/Code: Research Methods and Sample Survey Practice (Stat2032)

**Module title/code**: Sampling Theory and Research Methods (Stat-M2031)

Instructor: Abiyot Negash(MSc in Biostatistics)

**CHAPTER-1**

1. ***Research Methods***

1.1 Introduction: Concepts and definition of research

**What are research methods?**

Research methods are the strategies, processes or techniques utilized in the collection of data or evidence for analysis in order to uncover new information or create better understanding of a topic.

There are different types of research methods which use different tools for data collection.

**Qualitative Research** gathers data about lived experiences, emotions or behaviors, and the meanings individuals attach to them. It assists in enabling researchers to gain a better understanding of complex concepts, social interactions or cultural phenomena. This type of research is useful in the exploration of how or why things have occurred, interpreting events and describing actions

**Data collection tools**

Techniques or tools used for gathering research data include:

| **Qualitative Techniques or Tools** | **Quantitative Techniques or Tools** |
| --- | --- |
| **Interviews**: these can be structured, semi-structured or unstructured in-depth sessions with the researcher and a participant. | **Surveys or questionnaires**: which ask the same questions to large numbers of participants or use Likert scales which measure opinions as numerical data. |
| **Focus groups**: with several participants discussing a particular topic or a set of questions. Researchers can be facilitators or observers. | **Observation**: which can either involve counting the number of times a specific phenomenon occurs, or the coding of observational data in order to translate it into numbers. |
| **Observations**: On-site, in-context or role-play options. | **Document screening**: sourcing numerical data from financial reports or counting word occurrences. |
| **Document analysis**: Interrogation of correspondence (letters, diaries, emails etc) or reports. | **Experiments**: testing hypotheses in laboratories, testing cause and effect relationships, through field experiments, or via quasi- or natural experiments. |
| **Oral history or life stories** |
| : Remembrances or memories of experiences told to the researcher. |
|  |
|  |

**1.2 Dimension of Research**

The dimensions of research include the distinction between applied and basic research, or how people use research. The next is the purpose of doing research, or the goal of a study.  
One way to see the dimensions is as decision points for a researcher when moving from a broad topic to a focused research question to the design of a specific study.

|  |  |
| --- | --- |
| Dimension of Research | Major Types |
| How research is used | Basic, Applied |
| Purpose of the study | Exploratory, descriptive, explanatory |
| The way time enters in | Cross-sectional, longitudinal ( time series, panel, cohort), case study |
| Technique for collecting data |  |
| For Quantitative Data | Experiments, surveys, content analysis, existing statistics studies |
| For Qualitative | Field research, historical comparative research |

**The Use of Research**

Those who seek an understanding of the fundamental nature of social reality are engaged in basic research. Applied researchers, by contrast, primarily want to apply and tailor knowledge to address a specific practical issue. They want to answer a policy question or solve a pressing social problem.

**Basic Research**

* Basic research advances fundamental knowledge about the social world. It focuses on refuting or supporting theories that explain how the social world operates, what makes things happen, why social relations are a certain way, and why society changes. Basic research is the source of most new scientific ideas and ways of thinking about the world.
* It is true that knowledge produced by basic research often lacks practical applications in the short term. Yet, basic research provides a foundation for knowledge and understanding that are generalizable to many policy areas, problems, or areas of study.
* Basic research is the source of most of the tools- methods, theories, and ideas that applied researchers use.

**Applied Research**

* Applied researchers conduct a study to address a specific concern or to offer solutions to a problem of their employer, a club or organization they are affiliated with, their community or a social movement to which they are committed.
* Applied research means a quick, small scale study that provides practical results that people can use in the short term.
* The consumers of applies research findings are practitioners.

**Types of Applied Research**

**Evaluation Research**

* This type of research is widely used in large bureaucratic organizations to find out whether a program, a new way of doing something, a marketing campaign, a policy, and so forth is effective—in other words, “ does it work?”
* Most evaluation researcher adopt positivist approach and used cost benefit analysis.
* Two types of evaluation research are formative and summative. Formative evaluation is built in monitoring or continuous feedback on a program used for program management. Summative evaluation looks at final program outcomes. Both are usually necessary.

**Action Research**

* Action research is applied research that treats knowledge as a form of power and abolishes the line between research and social action.
* The researcher tries to advance a cause or improve conditions by expanding public awareness.
* AR are explicitly political, not value neutral.
* Action researchers assume that ordinary people can become aware of conditions and learn to take action that can bring about improvement.
* Action research is associated with the critical social science approach. It attracts researchers how hold specific perspectives.

**Social Impact Assessment research**

* Social Impact Assessment research may be part of a larger evnironemental impact statement require by government agencies.
* Its purpose is to estimate the likely consequence of a planned change.
* Such an assessment can be used for planning and making choices among alternative policies.

**Tools in Applied Research**

Applied researchers use two tools: needs assessment and cost benefit analysis.

**Need assessment**

In a need assessment, a researcher collects data to determine major needs and their severity. It is often a preliminary step before a government agency or charity deices on a strategy to help people.  
A researcher may confront dilemmas or difficult issues.

One issue is to decide on the group to target for the assessment. A second issue is that people may not express a need in a way that links it directly to policies or long term solutions.  
People often have multiple needs. A needs assessment may generate political controversy or suggest solutions beyond local control.

**Cost Benefit Analysis**

In cost benefit analysis, the researcher estimates the future costs and benefits of one or several proposed actions and gives them monetary values. Cost benefit analysis appears to be a neutral, rational, and technical decision making strategy, but it can be controversial. People do not necessary agree on what are positive and negative consequence. People look at cost / benefit differently. Someone’s cost maybe other’s benefits.

There are two ways to assign monetary value to costs and benefits.

Contingency evaluation asks people how much something is worth to them.  
Actual cost evaluation estimates the actual medical and job loss costs.

A significant issue for cost benefit analysis is the assumption that everything has a price and that people assign similar valuations.

**The purpose of a Study**

**Exploration**

* Exploration- the researcher’s goal is to formulate more precise questions that future research can answer. Exploratory research may be the first stage in sequence of studies.
* A researcher may need to conduct an exploratory study in order to know enough to design and execute a second, more systematic and extensive study.
* Exploratory research rarely yields definitive answers. It addresses the “what” question.
* It is difficult to conduct because there are few guidelines to follow.
* Everything is potentially important, steps are not well defined, and the direction of inquiry changes frequently.
* Exploratory researchers frequently use qualitative techniques for gathering data and they are less wedded to a specific theory or research question.

**Description**

* Description- the researcher presents a picture of the specific details of a situation, social setting, or relationship.
* Much of the social research found in scholarly journals or used for making policy decisions is descriptive.
* The researchers begin with a well defined subject and conducts research to describe it accurately. The outcome of a descriptive study is a detailed picture of the subject.
* Exploring new issues or explaining shy something happens is less of a concern for descriptive researchers than describing how things are.
* Descriptive research focuses on “how” and “who” questions. Descriptive researcher use most data gathering techniques—surveys, field research, content analysis, and historical comparative research.

**Explanation**

* When you encounter an issue that is already known and have a description of it, you might begin to wonder why things are the way they are.
* Explanation- the desire to know “why” is the purpose of explanatory research.
* It builds on exploratory and descriptive research and goes on to identify the reason something occurs.
* Explanatory research looks for causes and reasons.

**The Time Dimension in Research**

* An awareness of the time dimension will help researcher read or conduct research because different research questions or issues incorporate time in different ways.
* Quantitative research is divided into two groups: a single point in time (cross sectional research) versus multiple time points (longitudinal research). Quantitative research looks at a large group of cases, people, or units and measures a limited number of features.
* A case study involves qualitative methods and focuses on one or a few cases during a limited time period.
* Cross-Sectional Research
* In cross sectional research, researchers observe at one point in time.
* It is usually the simplest and least costly alternative. Its disadvantage is that it can not capture social processes or change.

**Longitudinal Research**

* Researcher examines features of people or other units at more than one time.
* It is usually more complex and costly than cross sectional research but it is more powerful, especially when researchers seek answer to questions about social change.

**3-types of longitudinal research are**

* Time series research is a longitudinal study in which the same type of information is collected on a group of people or other units across multiple time periods.
* The panel study observes exactly the same people, group, or organization across time periods.
* A cohort analysis observes a category of people who share a similar life experience in a specific time period.

**Case Studies**

* In case study research, the researcher examines, in depth, many features of a few cases over duration of time. A case can be individual, groups, organizations, movements, events, or geographic units.
* Qualitative and case study research are not identical, but “almost all qualitative research seek to construct representations based on in depth, detailed knowledge of cases”
* Case studies help researchers connect the micro level, or the actions of individual people, to the macro level, or large scale social structures and process.
* Case studies are likely to produce the best theory.

**Data Collection Technique Used**

The techniques are grouped into two categories: quantitative, collecting data in the form of numbers, and qualitative, collecting data in the form of word or pictures.

##### **Quantitative research tools**

Quantitative methods involve the collection and analysis of objective data, often in numerical form. The research design is determined prior to the start of data collection and is not flexible. The research process, interventions and data collection tools (e.g. questionnaires) are standardized to minimize or control possible bias.

##### **Qualitative Data**

**Experiments**

* Experimental research uses the logic and principles found in natural science research.
* Experiments can be conducted in laboratories or in real life.
* Experiments are most effective for explanatory research.
* In most experiments, the researcher divides the people being studied into two or more groups.
* He or she then treats both groups identically; except that one group but not the other is given a condition he or she is interested in: the treatment.

**Survey**

* Survey techniques are used in descriptive or explanatory research.
* A survey researcher asks people questions in written questionnaire or during an interview then record answers.

**Content Analysis**

Content analysis is a technique for examine information, or content, in written or symbolic material. Content analysis is used for exploratory and explanatory research but is most often used in descriptive research.

**Existing Statistics**

* In existing statistics research, a researcher locates a source of previously collected information, often the form of government reports or previously conducted survey. He or she then reorganizes or combines the information in new ways to address a research question.
* The existing quantitative information consists of stored survey or other data that a researcher reexamines using various statistical procedures. This is called secondary analysis research.
* Existing research is most frequently used for descriptive research.

##### **Qualitative research techniques and tools**

Qualitative research is generally used to explore values, attitudes, opinions, feelings and behaviours of individuals and understand how these affect the individuals in question. Researchers using qualitative methods are concerned with individuals’ perceptions of specific topics, issues or situations and the meanings they assign to their lives. This kind of research is important for generating theory, developing policy, improving educational practice, justifying change for a particular practice, and illuminating social issues. It may also be used to explain the results of a previous quantitative study or to prepare for the development of a quantitative study.

If your research team decides to use qualitative methods in your study, you will need to describe how qualitative methods will provide the information to help you address your research objectives and research question(s). For example, qualitative research may be appropriate because you intend to explore the values and behaviours of individuals in the study area in relation to a public health intervention, and to understand how these affect the phenomena in question. For example, why do some households have bed nets but do not use them? Or, why do individuals in a study area decline services from a specialized antenatal clinic? Qualitative methods can provide context, a deeper understanding of stakeholders’ needs and participants’ perspectives.

When collecting qualitative data, it is preferable to use more than one data collection method. Obtaining information on the same phenomena in a variety of ways allows the researcher to triangulate the data, adding rigour to the research. By nature, qualitative data collection is emergent and the design is intentionally flexible to enable the researcher investigate themes (findings) in more detail as they emerge. Qualitative methods use data collection methodologies such as interviewing, observation, discussions and review of documents (e.g. diaries, historical documents). The results of qualitative research are descriptive or explanatory rather than predictive, and are typically time-consuming to collect and analyze.

Unlike quantitative data collection, qualitative data collection can be more flexible allowing the research to incorporate emerging themes in the ongoing data collection. This allows the researcher to test and validate findings as they collect the data. For example, perhaps in one in-depth interview, the researcher learns that people do not attend the lymphatic filariasis mass drug administration because they use traditional medicines and therefore feel that they are already under treatment. The researcher may then add a related question to subsequent in-depth interviews to see how prevalent this phenomenon is in the study population.

##### **Pre-testing**

All study instruments (quantitative and qualitative) should be pre-tested to check the validity and reliability of data collection tools. Pre-testing allows the research team to check whether the research instructions and questions are clear, context specific, and that adequate time has been allowed to administer the questionnaire, etc. Pre-testing should be conducted from a comparable study population and environment. Since data management is critical to the success of the research, the data management team should be available during the discussion that follows the pre-test, in order to incorporate changes into the final design of the tool and facilitate the incorporation of appropriate checks into the data entry system. This stage includes designing the forms for recording measurements, developing programmes for data entry, management and analysis; and planning dummy tabulations to ensure the appropriate variables are collected.

**Qualitative Data**

* Field Research
* Most field researchers conduct case studies on a small group of people for some length of time.
* Historical Comparative Research
* Historical comparative research examines aspects of social life in a past historical era or across different cultures.
* This kind of research combines theory with data collection.

**1.3 Research Design**

A research design is simply the framework or plan for a study that is used as a guide in collecting and analyzing the data. It is a blueprint that is followed in completing a study. Research design is the blue print for collection measurement and analysis of data. Actually it is a map that is usually developed to guide the research. Definitions: “Research design is a master plan specifying the methods and procedures for collection and analyzing the needed information.” William Zikmund

“Research design is the plan, structure and strategy of investigation conceived so as to obtain answers to research questions and to control variance.” Kerlinger

Thus we can say that a research design is the arrangement of condition for collection and analysis of data in a manner that aims to generalize the findings of the sample on the population. Purpose of a Research Design: Research designs are used for the following purposes;

(i) To minimize the expenditure: Research design carries an important influence on the reliability of the results attained .It therefore provides a solid base for the whole research. This makes the research as effective as possible by providing maximum information with minimum spending of effort, money and time by preparing the advance plan of all about the research.

(ii) To facilitate the smooth scaling: Research design is needed because it facilitates the smooth scaling of the various research operations, thereby making research as efficient as possible yielding maximal information with minimal expenditure of effort, time and money.

(iii) To collect the relevant data and technique: Research design stands for advance planning of the methods to be adopted for collecting the relevant data and the techniques to be used in their analysis, keeping in view the objective of the research and the availability of staff time and money. Poor preparation of research design upset the entire project.

(iv) To provide blue print for plans: Research design is needed due to the fact that it allows for the smooth working of many research operations. It is like blue print which we need in advance to plan the methods to be adopted for collecting the relevant data and techniques to be used in its analysis for preparation of research project. Just as for better economical and attractive construction of a house need a blue print and a map of that, similarly we needs a blue print or a design for the smooth flow of operation of research.

(v) To provide an overview to other experts: A research design provides an overview of all the research process and with the help of the design we can take the help and views of experts 20 of that field .The design helps the investigator to organize his ideas , which helps to recognize and fix his faults.

(vi) To provide a direction: A research design provides a proper or particular direction to the other executives and others who are helping us into the process. The researcher studies available, literature and learns about new (alternative approaches.

**1.4 Stating a Research Problem**

The problems lie everywhere around us. They even lie at our door step and in our backyards. Human nature is so complicated, that a problem solved for one individual may still exist for another individual, a problem solved for one class/ school/teacher/ situation/ system/time etc., still remains a problem for another class/ school/ teacher/ situation/system/time or a problem solved for the time being may reappear with a lapse of time. We become habitual of living in the age of problems i.e. we are so much surrounded by the problem that we suffers from ,”problem blindness”. But in order to solve the problem or making research we need to delimit the problem. Selection of problem is not the first step in research but identification of the problem is the first step in research. Selection of problem is governed by reflective thinking. It is wrong to think that identification of a problem means to select a topic of a research or statement of the problem. A topic or statement of the problem and research problem are not the synonyms but they are inclusive. The problem concerns with the functioning of the broader area of field studied, whereas a topic or title or statement of the problem is the verbal statement of the problem. The topic is the definition of the problem which delimits or pin points the task of a researcher. It is the usual practice of the researcher that they select the topic of the study from different sources especially from research abstracts. They do not identify the problem, but a problem is made on the basis of the topic. It results that they have no active involvement in their research activities, whatever, they do, do mechanically.

Definitions of the Problem: The obstacles which hinder our path are regarded as problem. Different definitions of the problem are given below; “Problem is the obstacle in the path of satisfying our needs.” John Geoffery

“Problem is a question which is to be solved.” John. G. Tornsand

“To define a problem means to put a fence around it, to separate it by careful distinctions from like questions found in related situations of need.” Whitney “

A problem is a question proposed for a solution generally speaking a problem exists when there is a no available answer to same question.” J.C. Townsend

“A problem is an interrogative sentence or statement that asks: What relation exists between two or more variables?” F.N. Kerlinger

“To define a problem means to specify it in detail and with precision each question and subordinate question to be answered is to be specified, the limits of the investigation must be determined. Frequently, it is necessary to review previous studies in order to determine just what is to be done. Sometimes it is necessary to formulate the point of view or educational theory on which the investigation is to be based. If certain assumptions are made, they must be explicitly noted.” Monero and Engelhart Identification of a Research Problem: The following steps are to be followed in identifying a research problem;

**Step I** Determining the field of research in which a researcher is keen to do the research work. **Step II** The researcher should develop the mastery on the area or it should be the field of his specialization.

**Step III** He should review the research conducted in area to know the recent trend and studies are being conducted in the area.

**Step IV** On the basis of review, he should consider the priority field of the study.

**Step V** He should draw an analogy and insight in identifying a problem or employ his personal experience of the field in locating the problem. He may take help of supervisor or expert of the field.

**Step VI** He should pin point specific aspect of the problem which is to be investigated. The Sources of the Problem:

(i) The classroom, school, home, community and other agencies of education are obvious sources.

(ii) Social developments and technological changes are constantly bringing forth new problems and opportunities for research.

(iii) Record of previous research such specialized sources as the encyclopedias of educational, research abstracts, research bulletins, research reports, journals of researches, dissertations and many similar publications are rich sources of research problems.

(iv) Text book assignments, special assignments, reports and term papers will suggest additional areas of needed research.

(v) Discussions-Classroom discussions, seminars and exchange of ideas with faculty members and fellow scholars and students will suggest many stimulating problems to be solved, close professional relationships, academic discussions and constructive academic climate are especially advantageous opportunities.

(vi) Questioning attitude: A questioning attitude towards prevailing practices and research oriented academic experience will effectively promote problem awareness.

(vii) The most practical source of problem is to consult supervisor, experts of the field and most experienced persons of the field. They may suggest most significant problems of the area. He can discuss certain issues of the area to emerge a problem. Although research problems should not be assigned or they should not be proposed and allotted by a guide but consultation with the more experienced faculty member or research worker is a desirable practice. One of the most important functions of the research guide is to help the student clarify his thinking, achieve a sense of focus and develop a manageable problem from one that may be vague and too complex. Statement of Problem: Kerlinger has identified following three criteria of good problem statements;

1. A problem should be concerned with relation between two or more variables.

2. It should be stated ‘clearly and unambiguously in question form’.

3. It should be amenable to empirical testing. Meeting these criteria in his problem statement will result, in a clear and concise idea of what the researcher wants to do. This sets the state for further planning.

Objectives of Assumptions about the Problem:

1. To make the research work feasible.

2. To delimit the scope of the problem.

3. To establish the proper frame of reference.

Aspects of Delimiting a Problem:

1. Delimited to certain variables that should be mentioned clearly in the problem.

2. Delimited to the area or level as primary level, secondary level, college or university level.

3. Delimited to size of sample, considering the time, energy and money.

4. Delimited to the best method only.

5. Delimited to the best available tool for measuring the variable.

6. Delimited to the most appropriate techniques.

7. Other delimitations particular to a problem. As the above delimitations help the researcher for conducting the study, the findings of studies also confine to these delimitations. Evaluation of the Problem: When considering a problem a researcher is required to ask himself a series of questions about it. These are helpful in the evaluation of the problem on the basis of personal suitability of the researcher and social value of the problem.

Following questions must be answered affirmatively before the study is under Taken:

1. Is the Problem Researchable?

2. Is the Problem New?

3. Is the Problem Significant?

4. Is the Problem Feasible for the Particular Researcher? In order to be feasible, a problem should agree with the following:

a) Research competencies of the Researcher

b) Interest and enthusiasm of the Researcher

c) Financial consideration in the Project

d) Time requirement for the Project

e) Administrative considerations in the Project.

# 1.5 How to write a research Proposal

A research proposal describes **what** you will investigate, **why** it’s important, and **how** you will do the research. The format of a research proposal varies between fields, but most proposals should contain at least these elements:

* [Cover page](https://www.scribbr.com/research-process/research-proposal/#title-page)
* [Introduction](https://www.scribbr.com/research-process/research-proposal/#introduction)
* [Literature review](https://www.scribbr.com/research-process/research-proposal/#literature-review)
* [Research design](https://www.scribbr.com/research-process/research-proposal/#research-design-and-methods)
* [Reference list](https://www.scribbr.com/research-process/research-proposal/#reference-list-or-bibliography)

There may be some variation in how the sections are named or divided, but the overall goals are always the same. This article takes you through a basic research proposal template and explains what you need to include in each part.

## Table of contents

1. [Purpose of a research proposal](https://www.scribbr.com/research-process/research-proposal/#purpose-of-a-research-proposal)
2. [Title page](https://www.scribbr.com/research-process/research-proposal/#title-page)
3. [Introduction](https://www.scribbr.com/research-process/research-proposal/#introduction)
4. [Literature review](https://www.scribbr.com/research-process/research-proposal/#literature-review)
5. [Research design and methods](https://www.scribbr.com/research-process/research-proposal/#research-design-and-methods)
6. [Implications and contribution to knowledge](https://www.scribbr.com/research-process/research-proposal/#implications-and-contribution-to-knowledge)
7. [Reference list or bibliography](https://www.scribbr.com/research-process/research-proposal/#reference-list-or-bibliography)
8. [Research schedule](https://www.scribbr.com/research-process/research-proposal/#research-schedule)
9. [Budget](https://www.scribbr.com/research-process/research-proposal/#budget)
10. [Revisions and Proofreading](https://www.scribbr.com/research-process/research-proposal/#revisions-and-proofreading)

## Purpose of a research proposal

Academics often have to write research proposals to get funding for their projects. As a student, you might have to write a research proposal to get your thesis or dissertation plan approved. All research proposals are designed to persuade someone — such as a funding body, educational institution, or supervisor — that your project is worthwhile.

|  |  |
| --- | --- |
| **Research proposal aims** | |
| **Relevance** | Convince the reader that your project is interesting, original and important |
| **Context** | Show that you are familiar with the field, you understand the current state of research on the topic, and your ideas have a strong academic basis |
| **Approach** | Make a case for your methodology, showing that you have carefully thought about the data, tools and procedures you will need to conduct the research |
| **Feasibility** | Confirm that the project is possible within the practical constraints of the program, institution or funding |

### How long is a research proposal?

The length of a research proposal varies dramatically. A bachelor’s or master’s thesis proposal can be just a few pages, while proposals for PhD dissertations and research funding are often very long and detailed. Although you write it before you begin the research, the proposal’s structure usually looks like a shorter version of a [thesis or dissertation](https://www.scribbr.com/category/dissertation/) (but without the [results](https://www.scribbr.com/dissertation/results/) and [discussion](https://www.scribbr.com/dissertation/discussion/) sections).

## Title page

Like your dissertation or thesis, the proposal will usually have a [title page](https://www.scribbr.com/dissertation/title-page/) that includes:

* The proposed title of your project
* Your name
* Your supervisor’s name
* The institution and department

Check with the department or funding body to see if there are any specific formatting requirements.

### Abstract and table of contents

If your proposal is very long, you might also have to include an [abstract](https://www.scribbr.com/dissertation/abstract/) and a [table of contents](https://www.scribbr.com/dissertation/table-of-contents/) to help the reader navigate the document.

## Introduction

The first part of your proposal is the initial pitch for your project, so make sure it succinctly explains what you want to do and why. It should:

* Introduce the [topic](https://www.scribbr.com/research-process/dissertation-topics/)
* Give background and context
* Outline your [problem statement](https://www.scribbr.com/research-process/problem-statement/) and [research question(s)](https://www.scribbr.com/research-process/research-questions/)

Some important questions to guide your introduction include:

* Who has an interest in the topic (e.g. scientists, practitioners, policymakers, particular members of society)?
* How much is already known about the problem?
* What is missing from current knowledge?
* What new insights will your research contribute?
* Why is this research worth doing?

If your proposal is very long, you might include separate sections with more detailed information on the background and context, problem statement, aims and objectives, and importance of the research.

## Literature review

It’s important to show that you’re familiar with the most important research on your topic. A strong [literature review](https://www.scribbr.com/dissertation/literature-review/) convinces the reader that your project has a solid foundation in existing knowledge or theory. It also shows that you’re not simply repeating what other people have already done or said. In this section, aim to demonstrate exactly how your project will contribute to conversations in the field.

* Compare and contrast: what are the main theories, methods, debates and controversies?
* Be critical: what are the strengths and weaknesses of different approaches?
* Show how your research fits in: how will you build on, challenge, or synthesize the work of others?

If you’re not sure where to begin, read our guide on [how to write a literature review](https://www.scribbr.com/dissertation/literature-review/).

## Research design and methods

Following the literature review, it’s a good idea to restate your main [objectives](https://www.scribbr.com/research-process/problem-statement/#step-3-set-your-aims-and-objectives), bringing the focus back to your own project. The [research design](https://www.scribbr.com/research-process/research-design/) or methodology section should describe the overall approach and practical steps you will take to answer your research questions.

|  |  |
| --- | --- |
| **Methodology in a research proposal** | |
| **Research type** | * Will you do [qualitative or quantitative research](https://www.scribbr.com/methodology/qualitative-quantitative-research/)? * Will you collect original data or work with [primary or secondary sources](https://www.scribbr.com/citing-sources/primary-and-secondary-sources/)? * Is your research design descriptive, [correlation](https://www.scribbr.com/methodology/correlational-research/), or[experimental](https://www.scribbr.com/methodology/experimental-design/)? |
| **Sources** | * Exactly what or who will you study (e.g. high school students in New York; Scottish newspaper archives 1976-80)? * How will you select subjects or sources (e.g. [random sampling](https://www.scribbr.com/methodology/sampling-methods/), [case studies](https://www.scribbr.com/methodology/case-study/))? * When and where will you collect the data? |
| [**Research methods**](https://www.scribbr.com/category/methodology/) | * What tools and procedures will you use (e.g. [surveys](https://www.scribbr.com/methodology/survey-research/), interviews, observations, experiments) to collect and analyze data? * Why are these the best methods to answer your research questions? |
| **Practicalities** | * How much time will you need to collect the data? * How will you gain access to participants or sources? * Do you foresee any potential obstacles, and how will you address them? |

Make sure not to simply write a list of methods. Aim to make an argument for why this is the most appropriate, valid and reliable approach to answering your questions.

## Implications and contribution to knowledge

To finish your proposal on a strong note, you can explore the potential implications of the research for theory or practice, and emphasize again what you aim to contribute to existing knowledge on the topic. For example, your results might have implications for:

* Improving processes in a specific location or field
* Informing policy objectives
* Strengthening a theory or model
* Challenging popular or scientific assumptions
* Creating a basis for further research

## Reference list or bibliography

Your research proposal must include proper [citations](https://www.scribbr.com/category/citing-sources/) for every source you have used, and full publication details should always be included in the [reference list](https://www.scribbr.com/apa-style/apa-reference-page/). To create citations quickly and easily, you can use our free [APA citation generator](https://www.scribbr.com/apa-citation-generator/). In some cases, you might be asked to include a bibliography. This is a list of all the sources you consulted in preparing the proposal, even ones you did not cite in the text, and sometimes also other relevant sources that you plan to read. The aim is to show the full range of literature that will support your research project.

## Research schedule

In some cases, you might have to include a detailed timeline of the project, explaining exactly what you will do at each stage and how long it will take. Check the requirements of your program or funding body to see if this is required.

| **Example research schedule** | | |
| --- | --- | --- |
| **Research phase** | **Objectives** | **Deadline** |
| 1. Background research and literature review | * Meet with supervisor for initial discussion * Conduct a more extensive review of relevant literature * Refine the research questions * Develop a theoretical framework | 20th February |
| 2. Research design planning | * Design questionnaires * Identify online and offline channels for recruiting participants * Finalize [sampling methods](https://www.scribbr.com/methodology/sampling-methods/) and data analysis methods | 13th March |
| 3. Data collection and preparation | * Recruit participants and send out questionnaires * Conduct semi-structured interviews with selected participants * Transcribe and code interviews and clean [survey data](https://www.scribbr.com/methodology/survey-research/) | 24th April |
| 4. Data analysis | * Statistically analyze survey data * Conduct [thematic analysis](https://www.scribbr.com/methodology/thematic-analysis/) of interview transcripts * Draft the results and discussion chapters | 22nd May |
| 5. Writing | * Complete a full thesis draft * Meet with supervisor to discuss feedback and revisions | 17th July |
| 6. Revision | * Redraft based on feedback * Get supervisor approval for final draft * [Proofread](https://www.scribbr.com/proofreading-editing/) * Print, bind and submit | 28th August |

## Budget

If you are applying for research funding, you will probably also have to include a detailed budget that shows how much each part of the project will cost. Make sure to check what type of costs the funding body will agree to cover, and only include relevant items in your budget. For each item, include:

* **Cost**: exactly how much money do you need?
* **Justification**: why is this cost necessary to complete the research?
* **Source**: how did you calculate the amount?

To determine your budget, think about:

* **Travel costs**: do you need to go to specific locations to collect data? How will you get there, how long will you spend there, and what will you do there (e.g. interviews, archival research)?
* **Materials**: do you need access to any tools or technologies? Are there training or installation costs?
* **Assistance**: do you need to hire research assistants for the project? What will they do and how much will you pay them? Will you outsource any other tasks such as transcription?
* **Time**: do you need to take leave from regular duties such as teaching? How much will you need to cover the time spent on the research?

## Revisions and Proofreading

As in any other piece of academic writing, it’s essential to redraft, edit and proofread your research proposal before you submit it. If you have the opportunity, ask a supervisor or colleague for feedback. For the best chance of approval, you might want to consider using a professional [proofreading service](https://www.scribbr.com/proofreading-editing/) to get rid of language errors, check your proposal’s [structure](https://www.scribbr.com/proofreading-editing/#structure), and improve your academic style.

**Planning of Sample Survey**

In generally, the planning of a sample survey has three major steps.

**Survey Design and Preparation**

Setting objectives of survey

Sample design

Preparation of sampling frames

Decision on type of survey

Preparing survey instrument

Planning timetable

Survey budget proposal

Conducting pilot survey

**Data Collection**

Organization of field work

Locating respondents

Collecting information

**Survey Analysis**

Preparation for processing (data files and data structures data checking, coding, data entry. Etc)

Performing statistical analysis (descriptive and inferential statistics)

Presenting methods and findings in study report

**1.3. Classification of Data**

Statistical data may be classified into basic types: primary and secondary data. Primary (original) data refer to those which are collected to meet the specific problem needs at hand. Primary data are data collected by the immediate user(s) of the data expressly for the experiment or survey being conducted. It is this data that we will normally be referring to when we talk about “collecting data”. The nature and type of primary data required would depend largely on the study objectives and vary from one field to another.

Secondary data refer to already existing information, which has previously been collected and reported by some individual or organization for their own purposes, and at latter stage at least some of that data will come to be made available to other individuals and organizations, secondary data can be obtained rapidly and inexpensively. It may be of considerable value, although the exact value will depend upon the type of study being carried out.

**Uses of Secondary Data**

Sometimes data requirements totally met from available secondary sources in which case there is no need to execute survey and generate primary data. The general rule is to exhaust all possible means to explore secondary data before deciding to mount on a comprehensive plan primary data.

In particular secondary data may provide a context (geographic, temporal, social), validation for primary data, which allow us to assess the quality and consistency of the primary data may act as a substitute for primary data.

Administrative record is the main source of secondary data, but there are other various internal and external sources, like records, reports, books, periodicals, newspapers and academic studies. Apart from time saving and cost, secondary data is less subject to intentional bias and the only alternative to inaccessible information, which is impossible to gather through primary data approach.

**Types of Secondary Data**

One of the subtle skills of information management is to know what secondary data is available in a given field. Expert researchers develop a comprehensive knowledge of secondary data sources. Willingness to use such data appropriately is also a hallmark of good research. One can group secondary data sources into categories: official and unofficial. Official secondary data comprise all information collected, processed and made available by legally constituted organizations, primarily by government departments and statutory authorities. Unofficial secondary data comprises all other forms of secondary information sources.

**Unofficial Secondary Sources**

**Private Research Results**

Many corporations and private consultants generate potentially useful secondary data. The major problem is access: the commercial confidentiality that is usually thought to attach to such material means that they are rarely available for public scrutiny.

**Research Reports, Research Papers, Textbooks**

The whole gamut of academic research and publication forms a major body of secondary data sources; we might even see textbooks as ‘tertiary’ sources. The expectation that research will be freely published means that most research is made available, although not always accessible.

**Opinion Polls**

The process of collecting public opinion via surveys and questionnaires has been developed to a high level of sophistication in recent decades, and the results of such surveys, if made public, can be an important contextual source.

**Market Research**

Like opinion polls, most market research is carried out by private organisations on behalf of specific clients. Depending on the requirements of the client, the results of such surveys may or not be made public. The commercial orientation of most surveys tends to limit their general applicability.

**On-line Databases**

Increasingly, secondary source are being made available in electronic form. These data may range from bibliographic information on-line reference databases to census data.

**Official Secondary Sources**

Official Statistics abound all developed societies and some developing countries. Governments collect data about their operations, usually for valid reasons, such as for administrative purposes and efficient long-and short-term planning

**Government Organizations**

Individual government departments- both state and federal-collect statistics about their area of responsibility, and many are published in some form. This may be include an annual report’, and regular statistical analyses, or occasional reports. Similarly, statutory authorities are also required to report to federal or state parliament, but increasingly present data about their operations as a component of the public relations process. In general there are three approaches of data gathering, namely, administrative records, sample surveying and census.

**Data from Administrative Records**

Specially, in most countries, government bodies collect administrative data, which can be used for production of statistics needed for their own use and for incorporation in the system of official statistics. In developed countries, a great part of demographic and social statistics is derived from such data, for instance, statistic on vital events, education, health, criminality, transport and communication, etc. essential parts of economic statistic are also based on administrative data, for instance, foreign trade statistics and data on the production or sale of commodities subject to excise taxes.

**National Periodic Surveys**

The premier generator and collator of publicly available secondary data are the national statistical office. The federal statutory authority is responsible for the independent collection and analysis of statistics, particularly in the economic and social fields. The principal output is a large number of periodic survey, issued in various forms on a weekly, monthly, quarterly or annual basis (depending on the data). The following table is a summary of the major types of periodic data collected by the statistical offices.

Demography: population estimates and trends, vital statistics, migration

Social statistics and indicators: education, welfare, health (morbidity and mortality)

Labour and prices: labour force. Income, prices

Agriculture: livestock, crop, fishing, agricultural practices

Industry: manufacturing, mining, gas and electricity (production and usage), retail activity, tourism, building and construction.

**Census**

The national statistics office is also responsible for the largest regular exercise in undertaking population and housing census, or any other censuses. This provides such a wealth of social and economic data, and is so widely used as a source of background data in geography, sociology and economics, that all researchers should have a solid understanding of its organization and contents.

**Advantages and Disadvantage of Secondary Data**

it is important to note that primary and secondary data have a role in the data collection process, and that good research technique includes knowing when to use them, and from where secondary data can be obtained. One should properly assess the advantages and disadvantage of using such data.

**Advantages**

The main advantages of using secondary data are as follows

Secondary data is usually available more cheaply, the collection is generally significantly quicker and easier than collecting the same ‘from scratch’

Existing data are likely to be available in more convenient from, involving dial-up access rather than dust removal.

Secondary data can give us access to otherwise-unavailable organization, individuals or locations.

Secondary data allows the researcher to extend the ‘time base’ of their study providing data about the earlier state of the system being studied

The fact that secondary data are likely to be pre-processed eliminates the time-consuming analysis stage

**Disadvantages**

The main disadvantages of using secondary data are as follows:

The method by which secondary data were collected is often unsown to the user of the data. This means that we are forced to rely on the skills and integrity of the people who collected and analysed the data.

With secondary data we may have little or no direct knowledge of the processing methods employed, and we will rarely have access to the original raw data to check them, i.e., lack of accuracy

It quickly becomes outdated in an ever-changing environment

Differences in classification or measurement

**Types of Inquiry in Formal and Informal Surveys**

The two common types of inquiry for the gathering of statistics data on various topics are census and sample surveys

**Census:** this is an investigation that covers every individual unit in the population being studied. In other words census is characterized by four essential features:

Individual enumeration of all units

University within a defined territory

Simultaneity to express population with reference to point of time

Defined periodicity to assess changes of population

The based known examples are the national censuses of population and housing, agriculture and industrial. Which are conducted by many countries on regular basis. Since those these censuses aim at exhaustive coverage of all the units of the population of interest they are usually massive operations and are therefore conducted at regular intervals once in five or ten years. Although it aims at enumerating all units in the country, this goal is seldom fully achieved in practice. Units may be accidentally or erroneously omitted from a census for various reasons and certain well-defined categories of units may be deliberately excluded from the scope of the census. The existence of either omissions or exclusions does not mean a census is a survey.

A census can also relate to a much smaller and specific population. For example:

A population census of an individual village;

An investigation covering all extension workers in a district;

A study of all farmers receiving credit from an agriculture bank,

A total patients treated in a hospital

Census is likely to be appropriate for these examples because of two general circumstances. One is where basic information is needed for general planning purposes, so that it is necessary to gather data on a range of topics, from all sectors of the population with a good level of accuracy. The other is where the population under investigation is small and readily identifiable, but may be fairly heterogeneous, so that it is both feasible and advisable to include the whole of the population in the study

**Statistical Sample Survey:** this is an investigation which only part of the population is studied. This type of study is appropriate mainly when resources are not sufficient to cover the whole population or when it is not feasible to consider the whole population in terms of time and level of treatment. The information gathered through sample survey must be generalized to the population in order to make overall conclusion.

A sample survey can be of any size. It can be of large-scale operation such as a national demographic or agricultural survey or general planning purposes or it may be a very small-scale investigation, such as a sample survey of farmers in a project area. Usually the selection is made at random, so that the sample is known to be representative of the whole population.

In sample surveys it is useful to distinguish between the target population and the survey population. Target population is the population, defined at the planning stage, for which the results are expected. Survey population is the population actually covered the implementation. The difference could exist due to the exclusion of some units from a survey because of non-coverage and non-response.

**Rapid Methods:** this method is one of the informal surveys, which includes a variety of investigation techniques used to obtain rapid and sometimes qualitative information. Rapid methods are used in situations either when little time and resources are available and limited information is still useful, or when no data are initially available, and a quick preliminary inquiry is necessary to provide direction for further study. It is also feasible to use this method when other methods are not technically appropriate.

In most social sciences five commonly methods of rapid appraisal are identified; key informant interviews, focus group discussions, community group interviews, direct observation, and informal surveys.

These methods typically involve an investigator or team working in the study area, observing or measuring or interviewing the characteristics of interest. The observation may be direct or indirect, for example, quantification of the crop mixtures found in the area, of the incidence of crop failure due to pretests of drought. It may alternatively be by interviewing informants, either selected at random or chosen as being knowledgeable about the subject or study. Local leaders, prominent farmers, or the elderly of a community, are possible examples.

**Limitations of This Method:** though these methods are applied for different purposes, it is important to kwon the limitations of these methods. The main limitations are as indicated below.

The reliability and validity of the information generated can be questionable in many instances, due to factors such as the use of informal sampling, individual biases of the investigator/ interviewer, and the difficulty in recording, coding, and analyzing the qualitative data.

Qualitative methods do not generate quantitative data from which generalizations can be made for a whole population.

The general credibility of these methods is low compared to formal survey methods. These limitations should be weighed against the obvious strengths of rapid methods.

**Case Study:** this is an inquiry in which a small number of study units are investigated in great detail. Selection of units is not necessarily on a random basis. The focus of a case study is on the detailed structures, patterns of inter-relationships observed within each individual case included in the study.

An example might be an investigation of the allocation of rural women’s time to different activities, using a detailed time budget approach, in order to arrive at appropriate definitions and classifications of economics. Here a small number of women, selected to cover a range of types of household and non-farm activity patters, would be appropriate.

**Experimentation:** it is a controlled method of observation in which the value of one or more independent variables is changed in order to assess its causal effect on one or more dependent variables. That is, a stimulus is applied to a subject, and the effect observed. Experiments can be conducted in many setting such as laboratory experiments, which are conducted inartificial settings and field experiments, which are conducted in natural settings. A considerable body of theory has been development on the design and analysis of experiments. Here it is presented only to reveal that it is a type of inquiry or means of obtaining information.

**CHAPTER 2**

**SAMPLING FRAMES**

**2.1. Definition**

In its simplest form a sampling frame is a listing of the units from which the sample selection is to be made at any stage of sampling. The units in the frame may be either areas or units of objects covering the items investigated in a survey. The units in the frame may be large or small, areas, households, persons, farms, or any identifiable items, and are generally known as area frame or list frame.

Different definition of frames encompasses different concepts and integrating these concepts will give the following definition. The frame consists of materials, procedure, and devices that identify, distinguish, and allow access to the elements of the target population. The frame is composed of a finite set of units to which the probability sampling scheme is applied. Rules or mechanisms for linking the frame units to the population elements are an integral part of the frame. The frame also includes auxiliary information (measure of size, demographic information) used for special sampling techniques, such as stratification and probability proportional to size sample selections, or special estimation techniques, such as ration or regression estimation.

In multistage sampling the sampling units used at the first stage of sampling are called primary sampling units (PSUs). Those used at the final (ultimate) stage are called ultimate sampling units (USUs). In designs with three or more stages, units used for the intermediate stages are called secondary or second-stage sampling units (SSUs), third-stage sampling units, and so on. Therefore, for surveys with multistage sample designs, a frame is needed for each stage of selection. For example, for the three-stage design the sampling units for household survey are:

PSUs: Districts (Woredas)

SSUs: EA (kebeles)

USUs: housing units (households)

Any sampling frame used for the first stage of selection must cover the entire survey population (the designated PSUs). At subsequent stage of stages of selection, frames are needed only for the sample units selected at the preceding stage. In the above case a list of districts (woredas) would be needed for the first stage of sample selection. Lists of EAs (Kebeles) would be needed for the second stage, but only for the sample districts (woredas). For the final stage, lists of housing units (households) are required only for sample EAs (Kebeles). In this study the term secondary sampling frame will be used for frames that are developed specifically for the second and subsequent stage of sample selection.

**2.2. Basic Consideration in the Choice of Sampling Frames**

The choice of suitable frames for all stages of sample selection is a critical aspect of the design for survey. The population coverage, the stages of sampling, the stratification used, the process of selection itself-every aspect of design is influenced by the sampling frames. Key considerations in the choice of sampling frames, regardless of the stage of sampling for which they are used, include the following: Intended use, frame units, coverage, media, content, and additional information.

**Intended Uses:** Sampling frames are used for sample selection and for making estimates based on sample data. The choice of the sampling method to be used at each stage of selection is limited by the information available for each frame unit at that stage. If the information consists only of attributes (e.g., urban/rural classification, identification of higher-level units), it is necessary to use an equal probability selection method with a recent or without stratification. However, if quantitative information or measure of size (e.g., counts of persons or household from a recent census) is available for all or virtually all frame units, this information can be used in connection with sample selection or estimation, or both.

**Frame Units:** the frame units are the sampling units included in the frame. The kinds of units in frames used for surveys include:

*Area units* such as administrative subdivisions, census enumeration areas, land areas (segment), and others. Area units cover specified land areas with defined boundaries.

*Non-area units* include housing units, persons, nomadic tribes, institutions, construction camps, and other items, and these units must have a clear definition.

**Coverage:** The coverage objective of the frame of frames used for a survey is to provide access to all of the elementary units in the survey population and to do so in such a way that every one of those units has a known (or knowable) probability of selection in the sample for the survey. Access is achieved by sampling from the frames, usually through two or more stages of selection and by the use of rules of association that like elementary units to the units that were selected at the final stage of selection, i.e, the USUs.

**Media:** sampling frames may be stored either on print on electronic media. For a frame stared on electronic medium, it is relatively easy to produce a printout of the entre frame or any portion desired, and organize in any desired format.

**Content:** the frame contains a record for each frame unit. The only item that is absolutely indispensable is a unique identifier of each unit. If a unit is selected, the numerical identifier provides the means of access to the unit in order to perform subsequent sampling operation or to collect survey data: the numerical identifier will be linked with other identifiers, such as place names or addresses of housing units. Either in the frame itself or on maps or other auxiliary materials.

**Additional information:** There are a number of possible reasons for collecting additional information during the construction of a sampling frame. One accurse when the definition of the universe or of the sampling bit to be covered is rather complicated to apply under field conditions, and also classificatory information is gathered during the frame listing, and the final decisions as to which units are to be excluded or included can be made at a later stage. Another common reason is for the purposes of stratification and allocation in which the stratifying information must be gathered and recorded during the frame listing.

**2.3. Desirable Properties of Frames**

The properties can be grouped into three major categories: properties related to quality, those related to efficiency, and those related to cost.

**Quality Related Properties**

Quality related properties of a frame are those properties which make it possible to minimize non-sampling errors, especially coverage errors that might occur because of deficiencies in the frame. Desirable quality related properties are that the frame:

*Consists of well-defined units*, meaning that area units have recognized boundaries that are clearly delineated on various types of maps, and for non-area units a precise standard definition of the unit be established.

*Units have adequate identifiers*; usually frame units will have both unique numerical identifiers (primary identifiers), and the other identifiers, such as names and addresses (secondary identifiers). The units in the list must be traceable in the field.

*Must be complete*; the completeness of a sampling frame deals with the extent to which the intended coverage is actually achieved and the extent to which the desired information for each frame unit is included in the frame. If actually incompleteness and duplication exist in a frame it can create a problem and introduce bias to the survey estimates.

*Are up-to-date*; for frames that are to be used more than once, procedures must be developed for periodic updating to ensure that they are up-to-date for some are likely to change with time.

*Must have stable units*. If there is a choice with respect to the kinds of units to be used in a frame, it is preferable to choose the most stable kinds of units available, i.e., those that are least subject to change in number, definition and size.

**Efficiency-related Properties**

Efficiency-related properties of frames are those qualities that make possible and facilitate the use of efficient survey designs. Efficiency in this context refers to the relationship, between sampling error and the cost of producing survey estimates; the most efficient survey design is the one that produces the desired level of precision at the lowest possible cost.

Perhaps the most important of these properties is the inclusion of accurate and up-to-date supplemental information for each frame unit. Measure of size, such as population, number of households, number of agricultural holders, and other size of measure, are useful. Measure of size can be used in the following ways:

To construct sampling units

To form strata of units classified by size

To determine the allocation of sample PSUs to strata

To select units with probability proportionate to size (PSS)

As auxiliary variables for ration or regression estimates

**Cost-related Properties**

The preparation of sampling frames can be an expensive exercise. Low cost of frame development can best be achieved by relating the development, maintenance and updating of frames for censuses and household surveys as a single, integrated ongoing process. If two alternative frame sources would result in the same quality and efficiency, the one with the lower costs of development, use and maintenance would obviously be preferred, i.e., low cost of acquisition/preparation, low cost of use and low cost of maintenance. The choice of frame for a survey must be based on assessing the cost of using that frame and the total error of the survey estimates when that particular frame is used. Above all, the costs of frame preparation must be considered at the planning stage, and must be budgeted for. They are likely to be significant proportion of the total costs, and related to an element of the survey work which is of critical importance in determining the eventual quality of the survey result.

In summary the sampling frame plays a central role in the design of a sample survey. It determines how well a population is covered, affects the method of enumeration and influences the efficiency with which a sample can be designed. A frame becomes more valuable if it contains some supplementary information, which can be used to improve sampling and estimation procedures. The structure of the frame, the information it contains, and the quality of the information will determine the type of sample designs and estimation procedures that can be used in a survey. Sample frames lacking auxiliary information support simple sample designs are used for selecting the sample. A simple random sample may be selected, or if the list is large, a systematic sample or a systematic sample of clusters may be used.

Many sample designs use auxiliary data to produce more efficient samples. Complex sample designs that are more efficient than simple random sampling, such as those employing stratification, probability proportional to size sample selection, or special estimation techniques such as ration or regression estimators, require additional information beyond the identify of the target elements.

The sampling framemust be *accurate and free from defects*. It should be exhaustive (no units omitted), non-repetitive (each unit listed once only), current or fresh list must be available (p to date), the units should be clearly identifiable without ambiguity, and the units in the list must be traceable in the field.

**CHAPTER 3**

**SAMPLE DESIGN**

**3.1. Sampling Methods**

The general aim of all sampling methods is to obtain a sample that is representative of the target population. By this we mean that, as much as possible, the information derived from the sample survey is the same as we would find if we carried out a census of the target population, allowing for inevitable variation in the estimates due to imprecision.

When selecting a sampling method we need some minimal prior knowledge of the target population; with this and some reasonable assumptions we can estimate a sample size required to achieve a reasonable estimate, with acceptable precision and accuracy, of population characteristics. How we actually decide which sampling units will be chosen makes up the sampling method. Sampling methods can be categorized according to the approach that take to the probability of a particularly unit being included. Most sampling methods attempt to select units such that each has a definable probability of being chosen. Moreover, most of these methods also attempt to ensure that each unit has the same chance of being included as every other unit in the sample frame. All methods that adopt this general approach are called probability sampling methods.

The basis of probability sampling is the selection of sampling units to make up the sample based on defining the chance that each unit in the sample frame will be included. If we have 100 units in the frame, and we decide that we should have a sample size of 10, we can define the probability of each unit being selected as one in ten, or 0.1 (assuming each unit has the same chance). As we shall see next, there, there are various sampling methods that we can use to select the units.

It is important feature of probability sampling that each time we apply the same method to the same sample frame we will generate a different sample. For a finite population we can simple combinatorial arithmetic to calculate how many samples we can draw from a particular sample frame such that no two samples are identical. It turns out that from any population of N objects we can draw NCn different samples, each of which contains n sampling units. In fact, in probability sampling we are concerned with the probability of each sample being chosen, rather than with the probability of choosing individual units. If each sample is a likely to be selected as every other sample (assuming equal probabilities), then each sampling unit automatically has the same chance of being included as every other sampling unit.

**3.2. Choice of Sample Design**

A sample design is a joint effort of the survey statistician and other experts such as subject matter specialists, data users, and survey executing agency. Mostly statisticians require information from other experts in order to propose a sample design that will meet the required specification of the users at the lowest possible cost. Among few issues on which they should discuss and reach agreement may include objectives of the survey, variables to be measured, type of estimates required, levels of reliability and validity needed for the estimates and any restriction placed on survey with respect to timeliness and costs.

**Setting Objectives and Preliminary Investigation of the Survey:** the survey objectives should be clearly specified and precisely stated at the outset. Other issues related to the objectives and relevant to the survey must be assessed at the early stage of the design. Depending on the scope and topics of the survey it may cover the following.

A clear specification of the problem and the cause of the problem into a precise and definite statement; it clarifies exactly what is to be investigated and why it should be investigated.

Identification and definition of the population to be studied (target population and the kinds of units to be covered), and a description of the coverage, such as geographic area, branch of the economic or social group, or other classification of the population covered by the survey;

A clear specification of the desired information to be collected in statistical terms, i.e., to determine the data requirements.

The level of breakdowns by which the results are to be tabulated; regions, age groups, sexes, residences, and any other population covered by the classification.

The level of accuracy desired or the specification of tolerable errors: The accuracy of a survey estimate is generally taken to mean the closeness of the estimate to an exact or ‘true vale’, which is nearly always unknown. The error of a particular survey estimate is the difference between that estimate and the true value of the quantity being estimated. It arises sampling errors and non-sampling errors, which need due consideration at design stage. Since sample size determination requires the desired confidence level and margin of error, it is important to specify these factors by considering the cost and precision required.

The kinds of result expected and, the users as well as the uses of the data;

***Timeliness***-how soon are the results needed; the utility of survey results falls off gradually with the passage of time following the data collection stage of the survey. The rate at which utility declines over time depends on the content and objectives of the survey. For example, political polls, relating to specific election, revision of the minimum wage, monthly labor force survey-their results are needed very quickly. Users of survey data often press for timeliness at the expense of accuracy. Therefore, one has to produce timely data to facilitate their actual use and maintain the responsibility to produce accurate data.

**Sampling Plan**

There are different ways of designing a sample survey, but the idea of optimum design started with the sampling plan features such as ***selection process*** and ***estimation procedures***. The selection process deals with the preparation of sampling frames, sample size determination, choice of design to be used, and sample selection method. The estimation procedure involves the process for computing the sample statistics and calculating the reliability of these estimates. The purpose is to develop a sample design that would meet reliability requirements at the lowest possible cost, or alternatively, to produce the most reliable estimates for a fixed expenditure of resources.

**Selection Process:**

After making an assessment of survey objectives, the kinds to topic to be covered, description of coverage, reporting levels, and other issues as discussed above, the next step in selection process is to make a choice of design.

*Choice of design*: there are different designs of sample, which are likely to be appropriate for different types of survey, and in different circumstances. It varies from the simplest kind of sample survey (simple random sampling –SRS) to a more complex large-scale sample survey design (multi-stage sample design). In general, there are two approaches of sampling stages-single stage sample design (un-stratified/stratified) and multi-stage sample design (unstratified/ stratified).

**Single Stage Sample Design**

Unstratified single-stage sampling involves sampling techniques such as simple random sampling, systematic sampling, varying probability sampling (probability proportional to size-PPS), and cluster sampling. Estimates of sample size required to obtain measures with a given precision will often be determined for each design by considering the objective of the inquiry, and the permissible margin of error in the estimates.

Stratified single-stage sampling deals with stratified simple random sampling, and stratified varying probability sampling. In stratified sampling the population is sub-divided into a number of groups called strata, and sampling is carried out independently in each stratum. Then one can use a different selection schemes such as SRS, PPS, etc, in different strata. In determining the sample size and allocating samples to deferent strata, one should take into account the size of strata, total cost of survey, and variability between strata. The following are the most common designs used at a single-stage or at any sampling stages of a survey.

**Simple Random Sampling (SRS):** it is the simplest kind of sampling method. It requires as a sampling frame a list of sampling units-households, farmers, institutions, or whatever else is being used-in any convenient order. The items listed must be numbered in sequence, starting from one for the first item at the head of the list and continuing up to as many as there are items listed. A table of random numbers is needed to obtain a random selection of these numbers, and the items, which have been given the selected numbers that from the sample chosen for the survey. The used of random numbers ensures that the sample units are chosen, each number within the chosen range has an equal chance or probability of selection. Since each element in the sample frame is given one number each unit has an equal chance of selection for the sample. The sampling could be performed with or without replacement.

**Liner Systematic Sample (LSS):** it is operationally a convenient method of selecting a sample. In a systematic sample we decide the sample size n from a population of size N. in this case, however, the population has to be organized in some way, such that we choose a starting point along the sequence by selecting the rth unit from one ‘end’ of the sequence, where r is less than n, and is usually chosen between 1 and k randomly. We then take the rest of the sample by adding k to r, where k is an integer number equal to N/n. we do this repeatedly until we reach the end of the sequence.

Provided that certain conditions are met, an LSS can be treated just like a SRS for the purpose of analysis. The basic requirement of this method is that the list used as the sampling frame must not have any intrinsic regularity or periodicity of its own. It is therefore necessary to check whether the frame to be used has built-in regularity of this kind. It is therefore, perfectly acceptable for the frame to be used for an LSS to be ordered or ranked overall in some way. Ordering is indeed advantageous as this can have the effect of making the sample more efficient.

**Varying Probability Sampling:** This method utilizes the use of auxiliary variable such as measure of size in which the size varies from unit to unit. Using this measure of size the selection is easily performed with PPS. A list of units with their estimated size, say Mi, is required, and we cumulate the values against each unit. Then a predetermined sample size (n) will be selected by using SRS or systematic sampling. If SRS is to used, we select a random numbers between 1 and unit we obtain the required samples. To apply a systematic sampling we select a random number between 1 and I, where ÷n .If the selected random number is r, then the sampling number will be r, r+I,r+21,…,(n-1) I +r. note that the probability of selection a units is

**Cluster Sampling:** Clusters can be defined as sampling units containing several elements that occur in groups naturally or formed artificially. A cluster has listing units associated with it in which the units can be geographical, temporal, or spatial in nature. Thus, cluster sampling can be defined as any sampling plan that uses a frame consisting of clusters of listing units. In a single-stage sampling, we select a sample clusters and completely cover all units within selected clusters. Clusters can be selected by a variety of sampling techniques. For example, we can select a sample of clusters by SRS or by systematic sampling or sampling with PPS.

The important reasons for using cluster sampling are feasibility and economy. If the only sampling frames readily available for the target population are lists of clusters, then the only feasible method of sampling is cluster sampling. That is why for surveys of human population, to compile lists of households for the purpose of survey never seems feasible in terms of time and resources. Listing costs and traveling cots almost always lowest for cluster sampling.

The disadvantage of cluster sampling is that the standard errors of estimates obtained from this design are often high compared with those obtained from samples of the same number of listing units chosen by other sampling designs. Therefore, one can choose the sampling design that gives the lowest possible standard error at a specified cost or, conversely, the sampling design that yield, at the lowest cost, estimates having specified standard errors.

**Stratified Random Sampling:** on occasion we may suspect that the target population actually consists of a series of separate domains, each of which may have, on average, different values for the properties we are studying. Thus, there are various reasons for stratifications and one must investigate these issues in detail before resorting it. The reasons could be to increase precision, separate estimates may be required, administrative convenience, and the nature of the population may force to use it.

In these circumstances we should apply sampling methods that taken such domains (sub-population) into account and generates a stratified random sample. For this design one should consider that overall sample size determination and allocation of samples to different strata by taking into account the size strata, the total cost of survey, and variability of the strata.

**Sample Size Estimation**

The sample size for a survey must be decided upon at the planning stage, together with the sample design. If done properly, the correct estimation of sample size is a significant statistical exercise. The sample size required depends upon three factors-the level of precision required in the estimate, this requires specifying the acceptable margin of error and the confidence level; the level of variability of the variables to be estimated, which could be measured by the standard error or coefficient of variation; and the sample design to be used, in which different designs will produce different levels of precision for the same sample size, or conversely different sample sizes for the same level of precision.

The basis for calculating the size of samples is that there is a minimum sample size required for a given population to provide estimates with an acceptable level of precision. Any sample larger than minimum size (if chosen properly) should yield results no less precise, but no necessarily more precise, than the minimum sample. This means that, although we may choose to use a larger sample for other reasons, there is no statistical basis for thinking that it will provide better result. On the other hand, a sample size less than the minimum will almost certainly produce results with a lower level of precision. Again, there may be other external factors that make it necessary to use a sample below this minimum. If the sample is too small the estimate will be too imprecise, but if the sample is too large, there will be more work but no necessary increase in precision.

But remember that we are primarily interested in accuracy. For this both an appropriate sample size with the proper sampling technique is required. If the sampling process is carried out correctly, using an effective sample size, the sample will be representative and the estimates it generates will be useful. Another area that needs attention in sample size determination is that several variables are equally important in a particular survey and the precision requirements for each of this will then produce a different estimate of the sample size needed. In this case, one should make an assessment to come up with a single estimated sample size. These include.

Estimating the sample size of different variables

Balancing the required level of precision and the resources available for conducting the survey

Give due consideration to the likely trade-offs between sampling and non-sampling errors (large sample size reduces sampling error, but it may have the effect of increasing non-sampling error)

**Multi-stage Sample Design**

For studies of large or geographically dispersed populations it is more convenient to use a multi-stage sample design. It is particularly appropriate where a large-scale survey is to be conducted, and where for logistic and organizational reasons it is convenient for the sample to be grouped together in a more limited number of geographical areas, rather than being spread thinly and sampling is adopted in a number of situations.

Sampling frames may not be available for all the ultimate observational units in the entire population

A multi-stage sampling plan may be more convenient than a single-stage sample of the ultimate units, as the cost of surveying and supervising, in large-scale survey, can be very high due to travel, identification, contact, etc.

It can be a convenient means of reducing response errors and improving sampling efficiency by reducing the intra-class correlation coefficient observed in natural sampling units.

In *un stratified multi-stage sampling*, the sample is selected in stages, i.e., the population is divided into a number of PSUs, which are sampled; then the selected PSUs, are sub-divided into a number of smaller second-stage units, which are again sampled; the process is continued until the ultimate sampling units are reached.

For multi-stage simple random sampling, at each stage the selection design is SRS, which an equal selection probability for each stage. For example, for a two-stage simple random sample the selection method is SRS at first stage with equal probability (1/total PSUs) and srs at second-stage varying probability sampling with two-stage design, the selection methods could be probability proportional to size either at both sampling stages (PPS/PPS) or PPS at first stage and SRS at second stage (PPS/SRS), and similar procedure will be followed for more than two stages.

A basic principle of scientific sampling is that every sampling unit must have a known, positive probability of being selected. Where the probabilities are equal, the sample design is known as self-weighting, and the formulae for calculating estimates are relatively straightforward. Where the sample design is not self-weighting, then the data relating to different sample units have to be weighted.

***Self-weighting Design* in two-stage Sampling**

In multi-stage sampling, there are different sample designs to choice from. For example, for two-stage sample with constant sampling fraction, an appropriate design would be SRS or LSS at first stage for selection of PSUs and again or SRS or LSS at second-stage to select the second stage units, i.e., at both stages simple random sample is used (SRS/SRS).

**Example: L**et us assume that the two kebeles A and B were selected by SRS from a population of 100 kebeles. Assume that kebele A has 500 household and kebele B has 50 households. If the sampling fraction is to be 4%, then for kebele B with 50 households, the sample size would be 2 household and for kebele A with 500 households, the sample size would be 20 households. The probability of selecting an individual household can be calculated from the joint probability of selecting the kebele and then selecting the household within the kebele, i.e.

Accordingly, probability of selecting a household in kebele A would be:

Probability of selecting a household in kebele B would be:

Science the selection for both kebeles is in proportion to the size of the kebeles, no weighting is needed, and the design is self-weighting in which the estimation of means, totals, ratios and proportions is straightforward.

Another possible choice of sample design would be two-stage sampling with probability proportional to size, where selection of PSUs at first stage is with PPS and a fixed number of sample SSUs can be selected by SRS or LSS. Now consider the selection of kebeles at first stage and selection of households at second stage. The procedure for PSS is to make a list of the kebeles, in any order, together with the total number of households in each. Then, for each kebele in turn make a cumulative sum of household in the list next to each kebele. Then make a selection of required number of kebeles by using SRS or systematic sampling.

Example: Suppose we have the following list of kebeles together with their individual and cumulative numbers of households. We wish to select a total of 15 households from three kebeles. The first step is to select three kebeles using PPS sampling.

|  |  |  |  |
| --- | --- | --- | --- |
| Kebele | No of Households | Cumulative no of households | Random numbers |
| A | 224 | 224 |  |
| B | 573 | 797 |  |
| C | 1140 | 1937 | 1163 |
| D | 253 | 2190 |  |
| E | 720 | 2910 |  |
| F | 654 | 3564 |  |
| G | 310 | 3874 |  |
| H | 270 | 4144 | 4010 |
| I | 379 | 4523 |  |
| J | 411 | 4934 |  |
| K | 217 | 5151 | 5094 |
| L | 399 | 5550 |  |
| M | 281 | 5831 |  |

For the selection of sample, we need four-digit random numbers, because the total cumulative size, 5831 has four-digits. Using SRS and an appropriate random numbers table, the first random number is 4010, which is associated with the cumulative size of kebele H. the second random number is 1163 and related to kebele C. the third random number is 5094, which corresponds to kebele K. the final sample of three kebeles therefore consists of kebeles H, C, and K. a fixed sample of five households would then be taken in each kebele using SRS or LSS methods. The probability of selecting a household in each of the three kebeles is calculated as follows.The probability of selecting kebelee x probability of selecting household.

The selection probabilities are the same in each kebele, and hence the design is self-weighting. The appropriate estimation formula for this sample design will be used.

**Estimation Procedure:**

The estimation characteristics will be a major objective in surveys. Population estimates will be calculated from sample data and reported together with an indication of the precision of the estimate obtained from the sample variance, typical estimates are totals, means, ratios and proportions, in which their standard errors will be required to enable confidence levels to be placed on estimates, and tests of significance can e carried out.

Calculation of population estimates are derived from the type of sampling design used for the survey. Based on the type of sampling design estimates are raised from an estimate of a small sample to an estimate of the population by multiplying by the inverse of the sampling fraction. The more complicated the design in terms of number of stages, variation in sampling fraction and sampling with or without replacement the more complicated will be the algebra for calculations.

**CHAPTER 4**

**METHODS OF COLLECTING THE DATA**

**4.1. Time Dimension in Survey**

Two types of survey are classified according to the time of data collection: longitudinal surveys, and cross-sectional surveys. Longitudinal surveys gather information at different points in time in order to study changes over extended period of time. Three different designs are used in longitudinal survey: panel studies, trend studies, and cohort studies.

Panel studies are studies in which the same subjects are surveyed at different times over an extended period. The investigator observes exactly the same people, group, or organization across time periods. In a trend study, different people from the same general population are surveyed at different times. In a cohort study, a specific population is followed over a length of time. Cross-sectional surveys study a cross section (sample) of a population at a single point in time. It is usually the simplest and least costly alternative. Its disadvantage is that it cannot capture social process or change.

**4.2. Data gathering Techniques**

We distinguish three basically different methods of collecting data. These are:

Extraction of data record

Self-administered questionnaire

Direct investigation-measurement (observation) of the subject and interviewing (face-to-face, telephone

Our first step is to decide on which of these three methods to use. The data collection method will be determined by the objective of the survey, the nature of the items of information, the operational feasibility and cost.

**4.2.1. Extraction of data from Records**

It is usually possible to answer of the questions a survey is intended to cover from available data. For example, a mass of information about the population studied by social survey is available in historical documents, statistical reports, records of institutions and other sources; it is up to the survey to derive what help he/she can from it. However, one must first consider carefully its suitability for the purpose. One must critically assess population coverage, definition, how accurate is the information, and if sufficiently to date. In addition, government departments, institutions like hospitals, and professional institutes, and business firms possess a mass of information relating to individuals, but not generally available to the outside research since the information is complied quite for another purpose. There are some areas where information from records is the only available sources, like financial information, distributive trade, educational and transport data, in which close cooperation with the concerned institution is vital to have access to the required data. Information from records may serve as complement for analysis and can used as a base for preliminary investigation. Therefore, it is advisable to examine exhaustively what is available in records before launching any surveys.

**4.2.2. Self-administered Questionnaire**

Mail and self-administered questionnaire is a method of data collection in which researchers can give questionnaires with instructions directly to respondents or mail them to respondents who read instructions and questions, then record their answers and give it back or return it by mail again to data collecting agency.

This type of survey has many advantages, which include:

It is the cheapest can be conducted by a single researcher

A researcher can send questionnaires to a wide geographical area

The respondent can complete the questionnaire when it is convenient and can check personal records if necessary.

Mail questionnaires offer anonymity and avoid interviewer bias

They are very effective, and response rates may be high for a target population that is well educated or has a strong interest in the topic or the survey organization.

The disadvantages of this method may include

A low responder rate is the biggest problem; the process of retuning the questionnaire may be unnecessarily extended. The research can raise response rates by sending non-respondents reminder letters, but this adds to the time and cost of data collection.

A research cannot control the conditions under which a mail questionnaire is completed.

Researchers cannot visually observe the respondent’s reactions to questions, physical characteristics, or settings.

Mail questionnaire is not suitable for illiterate community.

Therefore, the use of this method is limited to predominantly literate society, as the method requires a clear understanding of the survey concepts through reading and writing. Because of this its use is limited to developed countries with high percentage of literacy.

**4.2.3. Direct Investigation-Measurement (Observation) and interviewing (face-to face, telephone)**

Measurement or observation of the subject and interviewing a respondent and obtaining the report on the matter are two approaches, which are by no means exclusive. It is very common indeed to find both being used in the same survey. Some topics can only be investigated by one or other approach, but many can be investigated using either, and in such cases it is necessary to assess which is more suitable in the circumstances of the particular study. Therefore, type of question and the nature and status of the topic will determine whether a required piece of information can be gathered by measurement or interview approach.

**Measurements or Observations**

Information on a topic can be gathered by measurement if it is physically measurable or observable. Common types of data collected by observation and measurement include;

Land area measurement

Crop output measurement

Anthropometrics measurements

Animal weight gain

Instrument recordings or readings (e.g. rainfall, temperature, etc)

Physical measurement or examination of people

Counts of human, animal and plant populations

Direct observations of work

Exchange activities (e.g. purchase and sale prices)

Data collection by measurement can be undertaken in several ways. Some of these are:

The direct measurement of a physical characteristic using an instrument

The observation of people engaged in an activity; and

Recording of relevant aspects of their activities

**Interviewing (Face-to-Face, telephone)**

Face to face interview is a social process that involves the interviewer and respondent. It is the process in which the interviewer meets the respondents, explains the purpose of the study, forwards a set of questions and records the answers. It is widely used in economic and social surveys. Information may be collected by interviews for various. It may be information which could be measured directly but would require too much time or too great a use of manpower or funds to do so, in which case probably less accurate interview methods is used instead. It may be information that cannot be directly observed or measured because they relate to the past. It may be information about the respondent’s own knowledge, opinions, perceptions or attitude.

**Some advantages of Face-to Face Interviews:**

Face-to face interviews have the highest response rate and permit the longest questionnaires.

Interviews control the sequence of questions and can use some probes.

Respondent is likely to answer all the questions alone. Interviewers also can observe the surroundings and can use nonverbal communication and visual aids

Well-trained interviewers can ask all types of questions including complex questions.

The disadvantages of this method may include the following:

Cost is high that is, recruiting, training travel, supervision, and personnel costs for interviews can be high. Interviewers bias is also high in this method

The appearance, tone of voice, question wording, and so forth of the interview may affect the respondent.

The use of telephone interviewing for social surveys has increased in developed countries substantially in recent years because of the high penetration of telephone. Its major advantages are lower cost and faster completion, which relatively high response rate. The phone permits the survey to reach people who would not open their doors to an interviewer, but who might be willing to talk on the telephone. There may be less interviewer bias social desirability bias than with personal interviews. The main disadvantage of this method is that there is less opportunity for establishing rapport with the respondent than in face-to face situation. Another disadvantage is that households without telephones and those with unlisted numbers are automatically excluded from the survey, which may bias results. Those who have phone number blocking may simply ignore calls from unfamiliar number of the survey. We have to note that the use of this method is unpopular and very limited in developing countries.

**CHAPTER 5**

**INSTRUMENTS OF DATA COLLECTION**

**5.1. Type of Instruments**

A data collection instrument is a document used for gathering and recording of data in a survey. Basically there are two types of instruments to collect data: *structured questionnaire* and *unstructured questionnaire*.

The first type of instrument, structured questionnaire used mostly in formal sample survey, is a formalized schedule or from and contains an assembly of carefully formulated set of questions for information gathering. In other words, a structured questionnaire is one of the instruments used in data collection and which contains written questions that people respond to directly on the questionnaire from itself, with or without the aid of an interviewer. In a structured questionnaire, all questions are prearranged in some specified order and the range of possible responses for each question is provided.

The second type is a checklist of topics (unstructured questionnaire) used, mostly in qualitative survey, when enquiries are not appropriate for structured questionnaires. An unstructured questionnaire contains mostly open-ended questions. This type of instrument is used in an informal or exploratory survey and designed in the form of survey guides, tally sheets, observational forms, field notes, outline of questions, etc.

Most questionnaires used in sample survey combine structured and unstructured questions. Since questionnaire is the main data collection instrument in formal sample survey, this chapter will discuss the issues involved in questionnaire design and other activities related to it.

**5.2. Principles of Designing Questionnaire**

All surveys involve presenting respondents with a series of questions to be answered. The questions may be simple single-item measures or complex multiple-item scales. In whatever form it exists, especially socioeconomic survey data are basically what people say to the investigator in response to a question.

One major contributory element in the process of normal sample survey for maintaining data quality is the questionnaire design. In this approach, questionnaires need to be structured and its design is critical because survey analysis depends on the completeness of the topics covered. A well-designed questionnaire will enable us to ask the respondents the same questions in the same way and their answers must be recorded and coded uniformly so that data can be aggregated across the sample.

Error-free data transfer requires clear, comprehensive questions, good enumeration, and clearly set out answers, much of this process depends on good questionnaire design. The form must provide for coding and subsequent data entry for processing. In this respect there are some questionnaire design principles, which links between interview and data processing.

Regarding the content one must include the minimum number of topics to meet the objectives. Because of resource and time constraints we should focus on items of direct and major interest and avoid collection of any non-essential information

Time for the interview is another factor that must be kept reasonable and this limits the number of questions.

The questions must be easy to for the respondents to understand and to answer accurately and clearly.

The questionnaire should be easy to use as an interview guide for the enumerator and as an instrument for recording answers.

The questionnaire should be self-contained which include identification of the enumerator, respondent, date of interview and any other reference information such as geographical identification and other field details.

It should be designed in such a way that recorded answers can easily be edited, coded and transferred onto a computer file for data processing tabulation and statistical analysis.

The flow, structure and length of questionnaire should encourage and keep the interest of the respondent.

Careful thought should be given to the quality of presentation material such as paper, the size of the sheets used, the clarity of printing and the spaces provided for recording answer.

The process of design is creative and one should develop strong preferences of particular styles of layout and phraseology since there is no signal prescription around which a from can be modeled. A typical sequence of activities to design a form would have the following pattern.

Draw up a list of question topics from a mixture of theoretical models, empirical information, research evidence and terms of reference for the study;

For each topic phrase the specific information required;

List them in a logical order, following either a chronological or sequential pattern;

Decide for each questions how to record the interview response;

Make a first draft layout on the style of paper to be used;

Test the design on model respondents;

Prepare a pilot draft for a pilot or test survey;

Modify the form the result of the test; and

Finalize the design and layout.

Review as many times as possible the number of questions finally listed.

From design is largely a compromise between opposing criteria: layout for collection versus layout for data processing. Layout for collection is the ease, speed and accuracy with which the questionnaire can be completed in the field, while layout for data processing is the ease, speed and accuracy with which information from the questionnaire can be processed for analysis. One should give equal attention to both aspects-collection and data processing.

**5.3. Type of Questions**

Two basic types of questions can be used questionnaires: open-ended questions and closed-ended questions depending on the amount of freedom given to respondent in offering responses. The type of questions for use will be determined by the form of responses sought the nature of the respondents and their ability to answer the questions.

**5.3.1. Closed-ended Question**

A closed-ended question is one where a predetermined list of alternate responses is presented to the respondent for checking the appropriate one(s). It implies that the respondents; answers are restricted in some way to a limited range of alternatives. Closed-ended question falls into one of two categories: dichotomous question and multiple-choice question.

A dichotomous question is one where a predetermined list of responses. Examples are yes-no, true-false, agree-disagree, like-dislike, fair-unfair and so on. A multiple-choice question offers more than two responses in the predetermined list of alternate responses. There are two categories of multiple-choice questions: single-coded question, where the respondent is permitted to check one and only one response; and multi-coded question, allows the respondent to select as many responses that are applicable.

Example:

Do you have a bank account? Yes =1, No = 2

How many children have you ever born? 1 = 1-2 2 = 3-4 3 = 5-6 4 = 7-8 5 = more than 8

Which type of soft drink(s) does your household consume?

1 = Pepsi-Cola 2 = Coca-Cola 3 = Mirinda 4 = Fanta 5 = Sprite 6= Seven-Up 7 = Others, Specify ---------------

Has the road construction activity had impact on your access to public services (health, education, market, etc)? Yes = 1, No = 2 if the answer is ‘yes’ explain the impact. ------

The choice can be made by making a mark alongside a category; by entering a numeric value; or by selecting a code from a code list. Setting categories of responses requires skill and experience in the areas of studies and suits computer processing.

The advantages of closed response categories are that:

It is easier and quicker for respondents to answer

The answers of different respondents are easier to standardized and to compare

The answers are easier to code and statistically analyze

The questions meaning is often made more by the response categories,

The answers are relatively complete as long as all relevant categories are specified

Respondents are more likely to answers about sensitive topics.

The disadvantages of closed response are that:

The respondent cannot guess at answers when they don’t know since have the categories to guide them

The appropriate category may be missing from the schedule

Failure to understand the question is less easily detected than with an open-ended question

A poorly planned list may act as a constraint to correct answers not catered for

Too few categories may fail to different between important groups, and enumerator error (placing the tick in the wrong box by accident will be more common)

**5.3.2. Open-ended Question**

An open-ended (unstructured, free response) question is one which allows the respondent to answer it freely in his or her own words, and to express any ideas generated from the question itself. Open implies that the respondent is permitted to answer in any form and at any length without any limitation on the range or complexity of the answer, to the question asked. Response categories are most often associated with exploratory or informal surveys, in which the investigator does not know the likely response from the units of study. It needs a checklist of topics, guidelines or unstructured questions.

For example: ‘which crops do you grow’? The question does not specify any particular season or crops or plots and hence many answers are possible. It is open for discussion. Why did you say you would not buy imported cooking oil when it is when it is available in the market? Again this could be discussed since the reasons could be quality, taste, price, etc.

The advantages of open-ended responses are:

The permit an unlimited number of possible answers which may not be consider at initial stage of the questions’ design

Respondent can answer in detail and can qualify and clarify responses by expressing in his/her own words.

Unanticipated findings can be discovered.

They permit creativity, self-expression, and richness of detail

They may be used when there are too many response categories to list on a questionnaire

They are useful when the questions are too complex to reduce to a few standard responses.

The disadvantages of open-ended responses are:

That much irrelevant information is collected

The answers are not standardized and are therefore difficult to compare and to make statistical analysis.

Coding responses is difficult

They require a higher level of skills on the part of the data collector since responses are written verbatim

More time, thought, and effort is necessary for completion

The forms are often bulky because answers take up a lot space in the questionnaire.

**5.4. Question Phrasing and Common Problems which arise with question phrasing**

Another aspect of questionnaire design that needs serious consideration is phrasing of the question. The information required should be well and clearly at each stage at which a questions is posed: initial definition and explanation in the survey manual; text in the questionnaire; precise units for physical measurement; and verbal phraseology by the enumerator. At each stage the question should have a clear meaning, the same meaning to every person asked and the researcher, an answer which the respondent knows, and an answer which can be given clearly and unambiguously by the respondent.

**Leading Questions**

A leading question is one that leads the respondent to choose one response over another by its wording. The presentation of questions should be neutral. The form of the question not indicate a preferred or ‘correct’ answer. For example the question, ‘you don’t smoke, do you?’ or ‘Do you buy the fertilizer recommended by the extension worker?’ leads respondents to state that they do not smoke in the first case, and that they should buy fertilizer recommended by the extension worker and that they are wrong if you fail to do so in the second case.

**Multiple Questions**

Multiple (soluble-barreled) questions are questions which combine two or more distinct questions into one single question. For example ‘Do you like listening radio and watching television?’ ‘Do you have a tractor or plough?’ “Does this company have pension and health insurance benefit?’ in this case one would be confused and undecided as to which answer one should offer. The best way to avoid confusion is to replace double questions with two or more single questions and then to ask only one question at a time.

**Ambiguous Question**

Ambiguity, confusion, and vagueness must be avoided from a question since different people will understand the question differently, and in effect their interpretation will depend on the indicial respondent. The question, ‘what is your income?’ could mean weekly, monthly, or annual; family or personal; from salary or from all sources; for this year or last year. The question, ‘do you drink beer frequently?’ is ambiguous because the word frequently does not specify a fixed time reference. Vague words and phrases like ‘kind of’, ‘fairly’, ‘generally’, ‘often’, regularly, etc., should be avoided.

**Probing Questions**

Probing is not easy. A delicate balance has to be struck between persistence and rudeness. Very often the respondent does not to tell the truth. In some culture it is socially acceptable to tell lies to close friends, never mind strangers. The enumerator working on a repeated visit survey has maintain a working relationship with the respondent and cannot permit the need to resolve minor contradiction on a few question to disrupt the relationship. In some cases unbelievable data have to be accepted, and it is helpful if some method is agreed for the enumerator to draw attention to this on the form.

**Use Simple Language**

The language of a question should be simple. The aim in the question wording is to communicate with respondents as nearly as possible in their own languages. Thus the wording of the question must be appropriate to the respondent. Question should avoid the use of technical terms and jargon, which the respondent may not understand. Where it is necessary to use technical or legal terms, one should provide definitions and explanations.

For example: ‘do you use inorganic fertilizer?’ it is better to specify types or brand names or colloquial terms with the respondent will be familiar. Also use terms which the respondent will understand and which will not cause offence. For example terms such as ‘peasant’, or ‘tribe’, or ‘witchdoctor’ may cause offence. It is important to conduct a pre-tests or pilot study to identify which common terms use and which ones may cause offence.

**Sensitive Topics**

In some cultures people do not like to discuss private matters openly. Sensitive questions are apt to be irritating, threatening, or embarrassing to the respondent. Such questions are prone to normative answer, answers which confirm that the respondent acts within the special rules of society even if that particular individual sometimes acts outside these rules. In a society which generally condemns drunkenness, question about drunkenness might generate denial even if drunkenness sometimes does occur. Under this circumstance it may be useful to word the questions so that there is some assumption that the activity does take place. Thus rather than ask do you ever get frank?’ we might ask ‘how often do you get drunk?’ the assumption in the question that you might sometimes get drunk may ease the guilt of the respondent and generate a more truthful answer.

Questions on age, physical or mental disability, deaths in households income, sexual behavior, family planning are relatively regarded as sensitive issues. Special attention should be given during field testing of the questionnaire to identify particularly tentative questions and how they can be improved by rewording or better interviewing procedure.

**6.5. Choice of the Reference Period**

During questionnaire design, the choice of appropriate time-reference period is an extremely important consideration. Time-reference period is the specified length of time for which the respondents is asked to give information about events occurring within it. The choice of reference period depends on the method of inquiry, the expected frequency of transaction, the account-keeping manner, and the recall lapse, in general, the more recent, and shorter a reference period, the better the information is likely to be.

**CHAPTER 6**

**PRE-TESTS AND PLOTSURVEY**

**6.1. Pre-Tests**

It is difficult to plan a survey without a good deal of knowledge of its subject matter, the population it is to cover, the way people will react to questions and even the possible answer they are likely to give. Particularly for large-scale survey it should be the general rule to conduct pretests and pilot survey in order to get solution to the following questions.

How is one to estimate how long the survey will take, how many interviews will be need, how much money it will cost?

How, without trial interviews, can one be sure that the questions will be as meaningful to the average respondent as to the survey expert?

How is one to decide which questions are worth asking at all?

Pretests and pilot surveys are standard practice with professional survey bodies and are widely used in research surveys.

The pretest is a preliminary application of the data gathering technique for the purpose of determining its adequacy. This may take the form of a series of small pre-test on isolated problems of the design. For example in testing of questionnaires, per-testing refers to one or more series of interview conducted on successive drafts of the questionnaire for the purpose of identifying and correcting errors and shortcomings. Its objective is to evaluate the general receptivity and feasibility of the questionnaire, and identify specific problems of communication between the interviewer and the respondent in terms of specific questions or items of information sought.

**6.2. Pilot Study**

A pilot survey or pilot study is generally a full-scale dress rehearsal of the survey. A major purpose of pilot study is to check whether the organization and arrangements of the survey actually work satisfactorily. The whole of the survey operation in all its aspects must be tested out on small scale. This approach thus checks the administrative and organizational arrangements in general, the arrangements for the supply and destruction of all the resources and equipment needed for the survey, as well as the fieldwork operations, the survey forms and manual, sample size determination and the data processing.

It should proceed through all the stages and operations of the survey proper, but on a small scale in a few selected localities. These localities should be chosen to cover as complete a range as possible to the types of area and population of different characteristics to be covered by the survey. But there may not be enough resources or time to cover as needed, in which case the priority is to cover a few areas but over a broad range of characteristics. Thus, the size and design of the pilot survey is a matter of convenience, time and money. It should be large enough to fulfill the above functions.

Since the purpose of the pilot study is to identify weaknesses and problems with the survey materials, procedures and arrangements, the senior technical staff should be closely involved in the pilot study to observe all stages of the work as it is being done under field conditions. In other word, the survey dorms and procedures must be observed under operational conditions in the field if problems are to be correctly indented, and appropriate solutions found.

If it is properly done, it is likely to lead to changes to the survey forms and manuals, and to the procedures and organizational arrangements. It is therefore necessary to allow enough time to analyze the result and observations from it, and produce revised materials and arrangements in good for the start of the main survey operations.

**6.3. Specific uses of pilot Survey**

The pilot survey has many benefits in particular if the survey is to be conducted for the time. In general it provides guidance on:

The adequacy of the sampling frame from which it is proposed to select the sample.

The estimates necessary for determining the size of sample needed in the actual survey so that the final estimates may be made with stated precision.

The non-response rate to be expected, i.e., the probable numbers of refusals and non-contacts can be roughly estimated from the pilot survey or pretests and ways of redacting non-response can be sought.

Making a sensible choice from alternative methods of collecting the data (observation, mail questionnaires, interviewers, etc.).

The adequacy of the questionnaire, which is probably the most valuable function of the pilot survey.

The efficiency of the instructions and general briefing of interviewers.

The codes chosen for pre-coded questions, which may help to decide the alternative answers to be allowed for in the coding.

The probable cost and duration of the main survey and of its various stages,

The deficiency of the organization in the field, in the office, and in the communication between the two.

**CHAPTER 7**

**SURVEY COST ESTIMATION**

**7.1. Time Scheduling**

Once there is an agreement to proceed with the survey, a planning timetable should be drawn up in order to facilitate planning and budgeting. Scheduling for field operations must take into account two key aspects.

List of survey activities; and

Approximate time needed to perform each activity

Important activities to be carried out starting from the beginning to the end must listed at the planning phase to ensure that certain activities are not overlooked. These activities should be listed against their target approximate time needed to perform each activity. Be realistic about the time necessary to complete each stage of the work. The following time schedule is an example of personal interview study.

|  |  |  |
| --- | --- | --- |
| No | **Activities** | **Time Needed** |
|  | Formulate survey objectives | 1 Week |
|  | Formulate survey methodology and design | 1 Week |
|  | Recruit and train interviewers | 3 Weeks |
|  | Draft pilot questionnaire | 2 Weeks |
|  | Pretest questionnaire | 2 Weeks |
|  | Finalize questionnaire | 3 Weeks |
|  | Compile sampling | 7 Weeks |
|  | Select the sample | 2 Weeks |
|  | Collect the data | 5 Weeks |
|  | Transfer data to machine-read media | 2 Weeks |
|  | Process the data | 3 Weeks |
|  | Prepare the report | 1 Week |
|  | **Total weeks** | **32 Weeks** |

Some of the activities are performed simultaneously while others need the completion of other activities and must be presented in a form of a chart so that the required time can easily be estimated. A bar chart approach in presenting the time schedule was developed by Henry L. Gantt and for the activities indicated above it shows a more realistic time span, in which 21 weeks is required instead of 32 weeks as illustrated below.

The following chart is an example of Gantt chart of study activities.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Important Activities** | **Time Required** | | | | | | | | | | |
|  | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| Formulate survey objectives |  |  |  |  |  |  |  |  |  |  |  |
| Formulate survey methodology and design |  |  |  |  |  |  |  |  |  |  |  |
| Recruit and train interviewers |  |  |  |  |  |  |  |  |  |  |  |
| Draft pilot questionnaire |  |  |  |  |  |  |  |  |  |  |  |
| Pretest questionnaire |  |  |  |  |  |  |  |  |  |  |  |
| Finalize questionnaire |  |  |  |  |  |  |  |  |  |  |  |
| Compile sampling frame |  |  |  |  |  |  |  |  |  |  |  |
| Select the sample |  |  |  |  |  |  |  |  |  |  |  |
| Collect the data |  |  |  |  |  |  |  |  |  |  |  |
| Transfer data to machine-read media |  |  |  |  |  |  |  |  |  |  |  |
| Process the data |  |  |  |  |  |  |  |  |  |  |  |
| Prepare the report |  |  |  |  |  |  |  |  |  |  |  |

**7.2. Preparing Budget**

Budget preparation in valves the assignments of cost to each survey activity. The main expenditure items include:

Office wages and salaries (administration, executive personal, quality control, data processing);

Survey materials;

Supervisory and interviewing costs (a numerations’, supervisors’ and field officers’ salaries and allowances);

Supplies for the reproduction of questionnaires, forms and manuals and other stationeries;

Transport cost;

Computer services;

Sampling design cost

Other administrative costs (office rentals, overheads recovery); etc

Preparation of a preliminary budget estimates is a priority activity that should be planned and executed at an early stage. The budget will depend on the survey design, including the levels of precision desired for various estimates, as well as on the geographical and other classification for the presentation of the results, and the operational conditions prevailing in the region.

**Example of Budget preparation for Survey**

|  |  |
| --- | --- |
| **Office Experts** |  |
| 1 survey director for 1 month at Bir 10,000 per month | 1,000 |
| 1 field organizer for 1 month at Bir 6000 per month | 6,000 |
| 1 survey statistician 1 month at Bir 6000 per month | 6000 |
| **Sub-total** | **22,000** |
| **Filed Personnel**  Salaries |  |
| 50 enumerators for 2 months at 400 Birr per month | 40000 |
| 10 field supervisors for 3 months at 6000 Birr per day | 18000 |
| 10 drivers for 3 months at 350 Birr per month | 10500 |
| **Sub-total** | **68500** |
| **Allowances** |  |
| 50 enumerator for 1.5 months at 25 Birr per day | 56250 |
| 10 field supervisions for 2 months at Birr per day | 18000 |
| 10 drivers for 2 months at 25 Birr per day | 15000 |
| 50 guides for 2 days at 10 Birr per day | 1000 |
| **Sub-total** | **90250** |
| Total field personnel | **158750** |
| Equipment and Transport |  |
| Office equipment and furniture | 80000 |
| Rent of vehicles | 150000 |
| Running costs, maintenance, insurance | 6000 |
| Enumerators’ equipment | 10000 |
| Data processing equipment | 4000 |
| Miscellaneous | 3000 |
| **Total equipment and supplies** | **253000** |
| **Stationary** |  |
| Printing of forms, questionnaires | 12000 |
| Pens, pencils, sharpeners, erasers, rulers | 500 |
| Report production | 1500 |
| Manuals | 2500 |
| **Total stationary** | **16500** |
| **Data processing staff** |  |
| 1 data expert for 2 months at 3000 Birr per month | 6000 |
| 5 data editors and coders for 20 days at 1500 Birr per month each | 5000 |
| 2 date entry clerks for 15 days at 2000 Birr month each | 2000 |
| Total data processing staff | 13000 |
| **Total Budget** | **463250** |
| **Contingency** |  |
| Approximately 10% | 46325 |
| **Grand Total** | **509575** |

**CHAPTER 8**

**FIELDWORK**

**8.1. Organization of Field Work**

After completing the preparatory activities, the actual fieldwork will be carried out. An investigator with responsibility for providing statistics for many subject matter areas must organize a fieldwork, because it is crucial, fundamental stage on which everything else depends. It involves recruiting and training of field staff, actual data collection, supervision, quality check, field administration and co-ordination. The survey organization must establish its policy on these and some of the common topic and, problems of fieldwork.

For formal surveys of relatively large scale it is more common to have to use lower-level filed workers. Filed workers are required for the collection of data where personal interview or measurement is used as a method of collecting data. The quality of these workers is one of the most crucial factors in determining the quality of information. It affects the quality of the data at the point they are collected, and errors at this point are the most difficult to detect, and the most difficult to correct or salvage at any later stage of the data collection process. Therefore, extra care and attention at the stage of recruitment of field workers is very essential.

One of the problems that are bond up with fieldwork is to employ full-time or part-time filed workers. There are three possible types of approach to the recruitment of field workers. Depending on the scale and type of survey:

Field workers might be recruited for a particular survey for a limited only; or

One should establish a permanent body of field workers to conduct a continuous survey program; or

Use an existing group of people, either from an established data collection organization of some kind such as a national (regional) statistical office, or a development agency or ministry, or a university, etc.

**8.2. Recruitment of Filed Workers**

The duties of interviewers include searching for obscure addresses, securing co-operation from respondents and administering questionnaires. Thus, they play a crucial role in performing the quality of field work. The value of the information obtained depends on his/her skill, good sense and accuracy. The attributes of “good” filed workers will vary depending on the cultural, physical environment and to extent on the content of the survey.

**Role of the Interviewers**

The primary role of an interviewer is to gather data upon which decision are based.

The interviewer must be well informed about the survey and its objectives.

The interviewer must establish good relations with the respondent, avoid arousing unnecessary prejudice, confusion or resentment, and always respect the confidence on which the respondent has given information.

The interviewers should make the respondent aware that all information collected will be treated as strictly confidential.

The interviewer must motivate the respondent to supply comprehensive and accurate answers.

**Selection and Qualities of Interviewers**

In selecting filed workers, in particular interviewers /enumerators, there is no single criterion to prescribe for recruitment. However, there is a considerable range of practices regarding the recruitment of enumerators, in which some of the recruitment points need to be considered in surveys. Some of these points are:

Establishing minimum requirements with resects to general education and physical fitness which include level of education, age limit, health conditions, language, sex, etc. for example, regarding age, persons outside 20 to 45 may not be appropriate for field work, and in some cases the upper age limit may go down to 35.

Testing enumerators on the ability to read maps and to make changes and clerical duties such as handwriting, form filling and ability to follow instruction;

Making of dummy interviews with field personnel and/ or observing enumerators working in the field; and

Assessment of interviewer qualities, which include the following being confident, appearing relaxed, being natural, conscientious regard for detail, absolute honesty and integrity, work under difficult conditions, abide by instructions, being able to write legibly and accurately, pleasant appearance and manner, tact (intelligence and education).

No matter how thorough and searching recruitment is, there will always be some chosen candidates who prove in practice to be unsuitable, or who find the work unsatisfactory. It is therefore necessary to make an allowance for wastage when making the initial selection, to include a period of trial employment upon first recruitment of enumerators.

The functions of the field staff mainly fall into three categories: data collection; supervision and quality control of data; and administrative or clerical duties and co-ordination. A duty statement must be prepared in detail according to these activities. To operate such functions, all field staff must have appropriate qualifications and be sufficiently trained.

**8.3. Training of Field Workers**

A large-scale survey requires hiring several field workers. Good training, adequate pay and good supervision are important for consistent high-quality performance. An effective training program is essential to complete field operations efficiently and on schedule. Before going into actual fieldwork, all the field workers should be trained on some important general interlocutory procedures and specific aspects of survey.

There are many satisfactory ways of framing training, but a few general principles can be applied. They should know how to give some insight into the overall work of the organization. This includes general issues as why the survey is being done, its relevance to the national and local development, the rationale of sampling (if applicable), the purpose of particular measurement techniques, by whom and how the results are to be used, about confidentiality of the information gathered and how to handle the survey in the field.

There are also several specific elements to field workers training. Such issues include, which usually indicated in a survey manual for regular use, description of the survey’s work, methods of data collection, interviewing techniques, how to check and handle completed questionnaires, what to do with non-response, standard definitions used in its questionnaires and content of questionnaires, for example, each individual question or item needs to be discussed, and issues of concepts, definitions, coverage, reference periods and inter-relationships between different questions all need to be fully explained.

The training procedure could be a formal training of courses that followed the teaching types of lecturing, demonstrations and discussion in the class, practice of mock interviewers in the class, trial interviews or practices in the field, discussions on the results from field practices and performing evaluation before deployment.

During the initial training period, survey manual is needed for immediate use and for reference purposes during subsequent fieldwork, and refresher training. It must be fully comprehensive and it must be easy to use, both to obtain a general grounding, and for referring to specific points. The full manual needs to cover all the topics that include the following.

Context of the survey at least should treat reasons for conduction it, relevance to development, purpose of particular questions and place within survey program.

Content of the survey may contain relationships between the questions, and for each question indicate the concepts, definitions, and coverage and reference period.

Techniques of data collection should include measurement techniques and interview techniques such as method of approaching respondents, phrasing questions, checking and probing.

Moreover, supervisory staff needs additional training in supervision. This should include training in the personnel management skills of motivating and leading enumerators, in the specific skills of supervising and checking field work, in organization and record keeping, and in the training of enumerators.

**8.4. Management of Fieldwork**

Logistics organization is a major element in the planning and organization of a survey. It requires advance preparation and ensuring of all necessary reserves, equipment and materials are available at the correct place and the right time. It is then necessary to plan and arrange for their procurement or rent.

Specific arrangements are very much dependent on local conditions. Arrangement of training centre and transport facility must be made in advance by considering the magnitude of survey team and the local transport conditions. A part from transport, the largest items of equipment needed are likely to be measuring instruments of various kinds. The need for such items is survey-specific, depends on the type of survey to be conducted.

Next, the most survey-specific equipment of all is the survey forms and questionnaires, and survey manuals and other documents. These must all be planned, designed, printed and distributed in good time. A document control system for the completed survey forms must also be established, and any related summary data forms must also be designed, produced and distributed.

In addition, the question of stationary supplies is easy to forget, but vital to remember. The supply may include pens, pencils, erasers and pencil sharpeners, clip-boards, bags, etc.

**Public relation:** it is important to publicize the survey by informing and involving responsible local government and administrative personnel and the traditional community hierarchy in developing countries. Their cooperation is essential in giving local credibility to the survey work. They can also helpful in providing facilities or giving information on local conditions and may even be obstructive if not kept fully informed. In addition, channels of communication through local organizations are also useful for publicity purposes. Local social and community groups, farmers’ associations and cooperatives may all be appropriate for communication. All local leaders and other prominent people, can be particularly useful in spreading information, and even assist in persuading reluctant respondents once their own cooperation is gained. For publicity, depending on the nature of the survey, some other channels such as radio, television, newspapers and magazines, posters and leaflets can be used in all locally important languages.

Finally, the population effected by the survey must be given information about the survey. It is also important to give them the opportunity to ask questions about survey, and not simply told.

**8.5. Supervision and Quality Checks**

Supervision and quality checks are part of field management. In fieldwork, supervision is an important duty aspect of supervisors. They should play a major role in controlling, coordinating, checking the quality of field work and leading field administration. The fieldwork of supervisors mostly consists of the following activities:

Allocating the work to enumerators

Distributing supplies to and collecting supplies from the field,

Facilitating transport and deployment

Monitoring the progress of the field work and taking remedial action, if necessary

Authorizing and making payments to interviewers,

Dealing with different technical and administrative queries,

Coordinating the full range of administrative support required at field level, and

Facilitating co-operation of local officials,

Checking the quality of work

Since interviewers are human beings, their work are liable to have mistakes. It is therefore advisable that supervisors should observe the enumerators at work, and check that are following correct procedures of data collection, whether enumerators made all the measurements or interviews as expected, whether the response rate is satisfactory at different levels, whether interviewers are asking questions, interpreting and recording answers in accordance with instructions. There are some ways of checking the quality of interviewing but the three major ones are observation of measurements or interviews at work, review and editing of completed questionnaire, and re-interviews for a sub-sample of assigned units, for large-scale survey, it may be necessary to prepare manuals for supervisors and office staff. Its content must include administration coordination, supervision and quality checks.

**Chapter-9**

**Survey Analysis**

**9.1. Data Processing**

Data collected using statistical techniques are in the form of number, and these numbers represent values of variables, which measure characteristics of subjects, respondents, or other cases. The numbers are in a raw form (raw data), on questionnaires, note pads, recording sheets, or paper. The raw data needs to be converted into a form suitable for analysis and interpretation; data processing is, therefore, the link between data collection and data analysis. This can be achieved through sequences of activities, which include editing, coding, entry, and tabulation.

**Editing**

It refers to checking and correction of data manually or by computer. The checking involves whether the information contained in the questionnaire is complete, recorded in the prescribed manner, accurate, internally consistent and from eligible respondent. Checking for completeness involves ensuring that no sections or pages of the questionnaire are missing and no answer to any question item is omitted. It is important to check whether the respondent refused to give an answer, the interviewer forgot to ask the question or record the answer, or question was not applicable to the respondent.

Checking for consistency can also contribute to data quality. This implies correction on the basis of logical and substantive criteria, internal consistency and other information available within the questionnaire. One must also try to check whether the answers are accurate. Inaccuracy may occur due to carelessness or to a conscious attempt to give misleading answers, and it may arise from either respondent or interviewer. An important area here is to spot possible interview bias in data entry, for example, when a fixed patter of responses consistently appears in the questionnaires. In addition, one must check for respondents’ eligibility, whether the appropriate respondents were contacted or not.

There are two kinds of editing namely, field editing and central editing. Field editing is to uncover errors in recording responses during the data collection stage. It allows access to the respondent for correction and additional information. The central editing is performed when the completed questionnaires are returned to the office. The objectives are to correct major errors such as those relating to questionnaire identification, and to prepare questionnaire for coding and data entry so as to minimize the possibility of error in these latter operations.

The modes of editing include manual and computer editing. Manual editing is performed by a group of editors, usually the field supervisors or trained editors. These editors are given a set of editing instructions specifying in detail the rule and guidelines to be followed in editing. For example, when an error is detected, the editors insert the correction alongside the original entry which should be never be erased. The problem with manual editing is that it is time-consuming and costly exercise. The editors also liable to make mistakes and there is no guarantee that all erroneous responses will be detected. Computer editing involves the use of computer facilities to detect inconsistencies in the questionnaires. It allows a large number of editing (cleaning and validation) instructions to be executed simultaneously, and hence speed and accuracy are achieved.

As a whole, the important thing to remember in editing, regardless of the approach taken, is that the objective is to present the truest picture of the universe represented by the survey and not to hide deficiencies in the data collection operation.

**Coding**

Coding refers to the process of identifying and assigning a numeric character symbol to questionnaire entries with the objective to prepare the data in a form suitable for entry into the computer. The coding procedure is a set of rules stating that certain numbers are assigned to variable attributes. Researchers being thinking about a coding procedure and code beak before they collect data; For example, a survey research pre-codes a structured questionnaire before collecting data. With unstructured questionnaire, the survey responses will need to be classified and the post-coding will start.

**Data Entry**

The data must be transferred from raw data forms into a format for computers. The aim is to store the data in a machine-readable format, and them to use it for cleaning, validation and statistical analysis. A subsequent transfer of parts or all of the data from one sheet or file to another is also possible. There are different ways of transferring data of which direct entry and optical scan sheet are the most common methods. Direct data entry is the more common and generally more appropriate approach in developing countries.

When transferring data from one medium to another it will give rise to a number of possible errors. These errors may occur because some data may be lost, some data may be repeated and the values of some data items may be changed. All are due to entry clerks and can be spotted, but only through the time-consuming process of checking the entered data against the original forms. Methods of checking include:

Either double entry approach, which involves entering every data item twice, independently by different people, and compare the two versions for inconsistencies.

Or a data list can be compared visually with the original forms. This requires people working in pairs, in which one reads the correct data from the survey form while the other checks entries in the listing.

In undertaking the data processing activities mentioned above, the manpower required (editors, data entry clerks, etc) must be recruited and trained.

**Tabulation**

**Tabulating** refers to simply counting the number of elements/ cases that fall into each coded category. Its primary objective is to organize data by groups so as to present information in a quantifiable and readily understandable format. Data tabulation may take the form of a simple tabulation or a cross-tabulation. Simple tabulation involves counting single variable, and presents an empirical distribution of the number of observation that fall into each category of response. For example, data on gender can be tabulated for male and female.

|  |  |
| --- | --- |
| Simple tabulation sex: |  |
| **Class** | **Frequency** |
| Male | 500 |
| Female | 450 |
| Total | **950** |

**Cross-tabulation** is a technique organizing data by specific group, categories or classes to facilitate comparison. In cross-tabulation, two or more of the variables are treated simultaneously; the numbers of cases that have the joint characteristic are counted. An example of three-way cross-classification is shown below.

Number of smokers by sex and age

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Smoking | | | | Sex and age | | | |
| Total | Male |  |  | Female |  |  |
| Total | 12-29 | 30+ | Total | 12-29 | 30+ |
| Total | 950 | 500 | 235 | 265 | 450 | 212 | 238 |
| Heavy smokers | 167 | 161 | 81 | 80 | 6 | 1 | 5 |
| Light smokers | 42 | 29 | 17 | 12 | 13 | 8 | 5 |
| Non-smokers | 741 | 310 | 137 | 173 | 431 | 208 | 228 |

Ideally, the general tabulation plan would have to be devised at the questionnaire design stage. However, at the stage of implementation, the data processor requires from the subject-matter specialist and unambiguous specification of exactly how each table is constructed and what its layout is. Before tabulation it would be necessary to decide whether to use the weighted or the weighted data. If weighted data are to be used, it is important to weight the data appropriately depending on the type of sampling design used, and to adjust for non-coverage and non-response. Another way of presenting the same tabulated data is in graphic form. Some common types of graphic presentations are the histogram, bar chart, pie chart, line graph, and frequency polygon.

**9.2. Analysis and Interpretation of Data**

Data analysis is concerned with categorizing, ordering and summarizing data while interpretation is essentially a follow-up which first involves the search for the meaning and applications of the survey results, and secondly makes references relative to the problem being studied, and ultimately draws conclusions about these relationships. The main divisions of data analysis and interpretation, together with the respective statistical tools and techniques adopted, are outlined as follows.

Describing Data

Measure of central tendency (arithmetic mean, median, mode)

Measure of dispersion (range, quartile deviation, mean deviation, standard deviation)

Statistical estimation (point and interval estimation, assessing differences)

**Testing Hypotheses**

Formulate the null and alternative hypotheses

Specify the level of significance (α)

Select the appropriate test statistics

Determine the critical value for the chosen level

Compute the value of the test statistic, using sample data

Compare computed value of test statistics and the critical value

Reject Ho if computed test statistic falls outside the acceptance region

Accept Ho if computed test statistic falls within the acceptance region

In significance testing, the large number of statistic techniques is available and one has to select the most appropriate one for use in a particular situation. The appropriate test will vary according to the scale level of data. *Statistical test for nominal data* include chi-square analysis (chi-square goodness-of-fit test, chi-square test of independence), McNemar test, Cochran Q test.

*Statistical tests for ordinal data* include kolmogoov-smirnov test, median test, mann-whitney U test, kruskal-wallis test, wilcoxon T-test, friedman two-way analysis-of-variance. *Statistical test for interval and ratio data*: the z-test and t-test are more powerful statistic test which are designed specifically for interval and ratio data.

**Measuring Association**

Many statistical techniques are available to measure the intensity of association between variables. This part requires relatively more sophisticated statistical techniques. The association can be measures by simple and multiple regression and correlation, multiple discrimination analysis, multivariate analysis of variance, canonical analysis, factor analysis, cluster analysis, dimensional scaling. The use of the appropriate techniques would depend on the number of variables involved, the level of measurement, whether variables are dichotomous or multichotomous.

**9.3. Presenting the Results**

The final phase of the survey is presentation of survey result in the form of a report. The primary purpose of writing a report is to convey information in a clear and concise manner to the reader. A good report is logically structures, directs the reader to those findings which convey information and draws firm conclusions and recommendations. It is the only one step the entire survey process which brings all other steps together and, in fact, documents all of the preceding efforts.

**Types of Survey Report**

A survey report may be technical report or popular report (non-technical report) in its nature, depending on the audience for which it is designed. The differ considerably in terms of detail, style of writing, use of technical terms and length. A technical report is generally interceded for other researchers or project managers who are interested in the technical details about research design, sampling design, statistical methods, tables, etc, adopted for the study. A technical report is free to use technical terms such as standard deviation, analysis of variance, degree of freedom, confidence level and the like. A popular report is intended for a more general audience who is interested in reading the survey findings but would not be particularly bothered about the survey techniques as opted. It makes less use of detailed, complex statistical tables. It is designed for rapid reading and easy comprehension of the main findings the survey. With these objectives in mind, the will normally make more use of flow diagrams, pictures, charts, and graphs.

**Structure of Survey Report**

There is no standard style of format for a survey report. The form, length,, style and degree of technicality of a survey report will depend on the subject, size of the study, type of reader for whom it is intended and, to a lesser extent, the relationship between the survey sponsor and the investigator. The major components of an actual report appear below, which is merely a suggested format.

Title page

Contents page

Executive summary

Background/introduction

Survey objective/presenting the problem

Survey methodology/describing the methods

Detailed findings/results and tables

Conclusion and recommendations

Appendix and references

**CHAPTER 10**

**Non-Sampling Error Survey**

**10.1. The Nature of Survey Error**

The ultimate objective of any statistical survey is to produce estimates for specified characteristics, applicable to specified population at a given time. The estimates are the results of the survey which may require a series of steps in the process. These results of a survey are used to make quantitative statements about the population studied. These may be descriptive statements about the aggregate population,

A survey error occurs when there is a discrepancy between the statements and righty. The error of a particular survey estimate is the difference between the estimate and the true of the survey.

Survey errors are generally divided into two major types

Sampling errors

Non-sampling errors

Sampling errors are present by design and result from the conscious choice to study a subject to get an estimate from a sample rather than from the whole population.

Sampling errors are not the result of mistakes persist, a through mistakes in judgment when designing a sample may carouse larger errors than necessary.

**Non-sampling errors** comprise all errors that contribute to survey error, except sampling error; it would arise when even if the whole population units were investigated.

They are often thought of as being the entirely to mistakes and deficiencies during the deferent, execution and analysis of survey producers.

They raise from wrongly concaved definitions, imperfections in the tabulation plans, failure to obtain response from all sample members and so on.

Some common types of non-sampling error are response error, measurement errors, recording and transcription errors, processing errors, selection bias, etc. they are not random in nature like sampling error.

It can be measured by

Non-sampling errors have the ff prosperities

Errors get bigger as the sample size increases and as the complexity of the sampling procedure increases.

Difficult to measure the size of the most components of non-sampling error without external information of some sort.

The non-sampling component of the total error is likely to be at least as large as the sampling component in a well-designed survey.

Non-sampling error is not captured by our formal for standard error of estimates.

**10.2. Classification of Non-Sampling Error**

One approach to classify non-sampling errors is by the stage of the survey in which they occur. The three major stages of survey are survey design and preparation, data collection, and data processing and analysis.

A second method of approaching non-sampling errors is on the basis of observational and non-observational errors.

Observational errors include questionnaire error, data processing error, and analysis (reporting) error, while non-observational error consists of interviewer error, respondent error, and coverage error.

**10.2.1. Non-Observational**

**2) Coverage errors:** include non-coverage and over converge of survey units.

Non-coverage is failure to include some units of observation, either directly or implicitly in the operation sampling frame. Non-coverage will lead to error in the sampling results if the missed units differ in characteristics from the units covered. Palate to cover all units will result in an undercount of the total population.

Over coverage or duplication is the inclusion of some units in the frame more than once, giving them a larger than intended chance of selection into the sample.

The sum of the absolution values of non-coverage and over-coverage error given gross coverage error.

Sources of coverage error

Frames: it is necessary that units at each stage of the sampling frame be exhaustive, non-overlapping and uniquely identification in the fled.

All elements in the target population are counted for

Each lower stage unit belying to one and only one unit it the next higher stage

Unambiguous description of each unit

Use of inappropriate sampling frames. The use of fixed area sampling frames is not generally applicable for nomadic, semi-nomadic or highly mobile populations.

Incorrect application of sampling procedures: when filed workers are asked to select the sample themselves, errors are more likely to occur and more difficult to control. The reasons behind these some of the tend to furor smaller households to keep their workload small

Others, with good intentions, may substitute larger neighboring households for small households.

Some field workers may favor more accessible, centrally, located units to those at the boundaries of the sample area.

Clearly, asking enumerators to select the sample is a bad practice and should be avoided.

**Interviewer error:** Interviewers can be a source of error by failing to put the question clearly, by influencing respondents to answer incorrectly, by mis-recording correct responses, cheating interviewers’ variability, and errors in respondent selection.

**Respondent error:** can be broadly classified into two categories

**Non-response error:** arises from failure to include a designated sampling unit, or population element such as households, or other units of observation, which have been selected for inclusion in the survey, or some of the data item that were to be collected.

**Response error:** occur in the data collection phase of a survey, and are distinguished from errors which occur in the data it is incorrect, it refers to response error. The fear of the “evil eye” is known in some cultures in reporting births, farm produce, etc.

There are two basic sources of responses errors, errors arising from respondents and interviewers.

**10.2.2. Observational**

a) **Questionnaire error**: include poor design, types of questions used excessively long questionnaire, inadequate interviewer instructions, or wring measurements/ attitudinal scales used.

b) **Data processing error**: many be coursed by error in editing data, in coding, in computer data entry, and in tabulation.

c) **Analysis (reporting) error** refers to the inappropriate statistical methods used in the interpretations of data.