مـركـز الإحـصـاء
STATISTICS CENTRE

# Statistical Survey Implementation Guide

Methodology and Quality Guides - Guide No. (8)

# Table of Contents

# Introduction

This guide includes the key principles and procedures for planning, designing, and implementing statistical sample surveys. It covers various technical issues and topics related to the concepts, definitions, and mechanisms followed in implementing both types of statistical surveys, namely, comprehensive and sample surveys.

Preparing this guide falls within the framework of the Statistics Centre – Abu Dhabi's efforts to document evidence related to statistical operations. It aims to enable the technical statisticians in Abu Dhabi and data users to access the detailed procedures of planning, designing and implementing all types of surveys; economic, social statistics, etc.

This guide consists of eight chapters that follow The Generic Statistical Business Processing Model (GSBPM). The first chapter includes an introduction to statistical surveys. The second chapter shows the methods for formulating statistical surveys and the implementation plan. The third chapter introduces the procedures of designing statistical questionnaires. The fourth one present the key concepts for the design and selection of statistical samples. The fifth chapter presents the statistical data field collection processes. The sixth one relates to the key concepts of statistical data processing.

The seventh chapter tackles the statistical estimates, and the eighth one describes to disseminating statistical data.

# 1 | Introduction to Sample Surveys

## 1. Introduction

The statistical survey is defined as an organized statistical work that relies on a scientific basis and aims at providing statistical data on specific features of a statistical population. It is based on the principle of including all units of the population or part of it, whether by selecting a sample of the population through the method of probabilistic sampling or by including all population units, which is known as the comprehensive survey or census.

In accordance with the aforementioned, the implementation of a statistical survey arises from the need for information on special features in a statistical population that currently are either unavailable or insufficient. Usually, the entity requesting the survey targets the analysis of an economic, social, or demographic phenomenon or issue, among others, whereby this entity may be either a statistical entity or a governmental institution in Abu Dhabi, which might be an official institution or a private one.

## 2. Stages of Implementing a Statistical Survey

Obtaining accurate results which would reflect the status of the tackled phenomenon would require implementing a statistical survey according to a series of clear stages, where each of them includes a specific and defined set of formats and procedures. Accordingly, the stages are , design and build, data collection and editing, processing and analysis, and dissemination .

## 3. Characteristics and limitations of the Sample Surveys

**Characteristics of Sample Surveys**

Sample surveys cover the various needs about the detailed data related to many fields, including the economic and social areas, whereby such samples are implemented for many reasons, such as the following:

- When the survey aims at obtaining highly accurate data while incurring specific costs or achieving accuracy levels upon relying on the lowest number possible of the individuals of a population (while incurring the lowest costs possible);
- When the number of the individuals of a population is entirely unknown or when the census frame is wide, such as in terms of statistics about natural resources and minerals present in the earth core, among others;
- When the number of the individuals of a population is very big, and a comprehensive census is impossible in this respect;
- When time, money, and other resources are limited;
- When tackling many phenomena in terms of which a comprehensive census would cause losing some population units, the thing that would require conducting a sample survey to preserve the population.

**Limitations of Sample Surveys**

- Implementing a sample survey does not cover the population as a whole and does not produce the same results and conclusions drawn from a comprehensive survey.
- The implementation of sampling theories requires qualified and sufficiently trained individuals.
- The sample survey shall be carefully planned and implemented as any sampling error would expand whenever the results are multiplied by the sampling weight of the sample estimates..

## 4. Factors to be considered upon designing the sample survey

1. Available Technical Capabilities: This includes the locally available expertise and the ability to use time and effort-efficient software to process samples.

2. Available Financial Capabilities: Such capabilities are considered a factor for determining the sample size and methodology, given no survey shall be implemented in case it requires more than the available financial resources and staff. Hence, a balance shall be established between the suggested survey design and the available capabilities.

3. Dissemination Levels: They shall be considered in terms of the survey design. For instance, in case obtaining data and conclusions pertaining to detail geographical levels, such as a sector or a residential area, was required, the comprehensive census method shall be adopted. However, in case the required dissemination levels were higher, such as the case of a survey that must be implemented at the country or regional level, the sample survey may be adopted.

4. Available Time for Obtaining Results: The survey design shall meet its objectives, whereby the results shall be drawn as per the set time for such retrieval, whereby information is necessary within a specific timeframe, and it loses such importance and value in case its dissemination was delayed.

5. Required Accuracy Level: The most efficient designs achieve the highest accuracy level while incurring the lowest cost possible. The sample size is based upon the level of accuracy needed and the adopted survey design. Hence, once the survey size increases, the accuracy of the design increases as well.

6. Survey Objectives: It has been noted that the survey design aims at achieving a sole objective that is different from those achieved by a multi-objective survey design. Hence, the aforementioned would result in a difference in terms of the sample, work, and databases' size, among other survey variables.

7. Expected Non-Response: If a high rate of non-response was expectd, a sample survey should be adopted whereby the selected sample size shall be carefully adjusted upon taking the non-response rates into account.

8. Survey Periodicity: In case the survey was implemented annually, quarterly, or monthly, and if a comparison was to be undertaken between the years, quarters, or months, the survey design shall be different from the survey carried out once only.

## 5. Survey Population and Target Population

One of the stages of sample survey design and implementation relates to the identification of the target population, in other terms, the population units which the sampling unit will cover. Hence, a target population is defined as all the population units which are covered by the survey framework. For instance, in terms of the Household Income and Expenditures Survey, the target population would cover all the individuals who have a usual residence in the Emirate of Abu Dhabi.

The survey population is another concept that also relates to statistical surveys, whereby the survey population is defined as a partial group of the target population. The survey objectives may not cover the entire target population, whereby some population categories may be excluded in terms of many surveys. In this case, upon the exclusion of specific parts or units, the target population is called a survey population. For instance, in terms of the Household Income and Expenditures Survey, the target population would cover all the individuals who have a usual residence in the Emirate of Abu Dhabi. However, some of the individuals residing at labor camps may not be covered by the survey under specific circumstances. In this case, the survey population would encompass the residents upon excluding the residents of labor camps.

## 6. Sampling Frame and Sampling Units

A sample frame is defined as the list including all the population units, which may include infromation about the unit like address and location, or a map identifying the population units. It is a database from which the samples are withdrawn. Accordingly, the sampling frame is considered as the main element for implementing a sample survey.

Identifying a sample frame of a sample survey is based upon the definition of the statistical sampling unit adopted for such survey and for which statistical data is collected. On the other hand, the elaboration of a sample frame is based upon the availability of a population units' list. There are two primary sources for the development of a statistical sample frame.

**First:** Comprehensive census and surveys: including a comprehensive identification of all the population units, which would ensure granting an opportunity or probability for each of the population units to be appear in the sample. The population Census is considered as one of the primary sources for providing a sample frame for the households and social surveys. The same applies to the general census of economic establishments, which provides key data for elaborating economic establishments' sample frame.

**Second:** Administrative records data lists various economic, social, and demographic variables, among others. These lists may be obtained from competent governmental institutions and entities, and they may be adopted as a frame for a statistical sampling upon examining the efficiency, accuracy, and coverage levels of such data.

It is known that one of the main requirements for ensuring the validity of a statistical sample frame is their contemporarily, whereby such a factor reflects the status quo of the statistical population upon withdrawing the sample. Hence, a continuous update of the frame sample's input is necessary.

## 7. Statistical Survey Errors

The implementation of a statistical survey, whether a sample or a comprehensive survey, may feature statistical errors which may affect the survey results or data. Hence, mitigating such statistical surveys would require obtaining highly accurate and efficient data.

Statistical data collected using comprehensive surveys (census) or sample surveys may contain several errors. Practical experience has managed to classify many of these errors to make them known to those working in the statistical field, whereby such errors include, for instance, non-response errors, coverage errors, measurement errors, etc.

However, the sampling method also features a unique kind of error emanating from the sample's non-inclusion of a specific part of the population, rather than encompassing the entire population, as the case is in terms of the census. Hence, whether the sample is large or small, its data will include this kind of error.

Accordingly, this kind of error is called a sampling error to distinguish it from other types of errors that may be featured under census or comprehensive surveys (census), including non-sampling errors. In other terms, if we are to exclude all non-sampling errors, it shall be noted that sampling errors will undoubtedly occur. The chapter on the sampling techniques will tackle the details of sampling errors .

As for non-sampling errors, arise from any source other than those resulting from using the sampling method rather than a survey of all population units. These errors are considered extremely important because of the impossibility of calculating their size and their impact.

These errors include:

1. Errors resulting from a failure to design the form appropriately;
2. Errors arising in the data collection stage;
3. Lack of overall coverage, thus covering only a part of the population rather than all of it;
4. Errors resulting from the non-response of some of the selected units;
5. Errors of data processing and tabulation and results retrieval and printing.

# 2 | Objectives of the Statistical Survey and Implementation Plan

## 1. Procedures to Identify the Objectives of the Survey

Identifying the objectives of the statistical survey is one of the key stages for the implementation of a survey. Hence, once the objectives are set clearly, planning for the later stages of the survey would be easier, given that the implementation of these later steps is done in such a way aimed at ensuring the consistency of the final results of the survey with the main objectives that have been set.

However, it is also essential to continuously review and monitor the objectives and progress of the survey objectives throughout the various stages of implementation to ensure the consistency of these stages with the pre-set objectives and avoid deviating from the overall objective and the detailed objectives of the survey.

The identification of the objectives of the survey arises from studying a specific phenomenon or issue that might be economic, social, or demographic, or any other field that subsequently leads to the need to perform a survey to obtain data to reveal the relevant detail about such phenomena, as well as study the causes and providing solutions to it. Hence, accurate knowledge of the survey's objectives enables the statistician to identify the target population for the survey. If there is ambiguity surrounding the main purpose and the detailed objectives of the survey, it will hinder the ability to identify the unit of the population to be surveyed. For example, if the survey's objective on the labor force is ambiguous as to whether it targets studying unemployment among locals-only or includes locals and residents, it will not be possible to determine the target population.

Hence, identifying the statistical survey objectives includes several procedures or steps and requires coordinated efforts from all parties involved in the survey, including the executive parties, respondent, and user. The key procedures or steps followed in identifying the objectives are the following:

### A. Identifying the Variables and Needed Information

The primary step in identifying the survey's objectives in determining variables and the information needed on a specific phenomenon. This is done by setting a framework for the issue or phenomenon targeted by the survey and providing answers to the questions on the nature and the situations surrounding this issue. An example of such questions and inquiries what are the detailed information required from the respondent?, and what are the specific topics to be covered?. Where answers have been provided to all questions raised, these answers should be accredited as reliable information to be referred to at every step of the survey to ensure the achievement of the survey's objectives.

### B. Identifying Data Users and Benefits of Using these Data

Upon identifying objectives, it is necessary to identify the users of the data which will be collected during the survey as well as the possible uses of such data, and the expected results which would arise from using data, as the identification of users would enable determining their orientations, and hence include them in the objectives to be formulated while planning the statistical survey.

In addition to those mentioned above, it is also necessary to specify the reasons behind using the data, including whether such information is being used for describing a specific phenomenon or analyzing the relationship between the variables of a study, as well as to determine the decisions to be taken according to the survey data and the consequences of such decisions.

It is worth noting that respondents should be engaged and consulted to determine the objectives, identify the variables that lead to the achievement of these objectives, and specify and formulate the questions that reflect such variables and could be included in a statistical questionnaire.

## C. Identifying Key Concepts and the Target Population

To identify the requested data for achieving the survey objectives, it is necessary to define the terms and concepts used in the survey, whereby exceptions and survey technical information are identified accordingly. Moreover, it is also essential to use internationally adopted terms and definitions to facilitate comparisons and consistency in the survey. When using a new definition or a term that was not internationally adopted, an internal description for such a term should be provided and adopted by the statistical entity.

It is necessary to provide a professional and clear description for these statistical concepts, definitions and terms as well as to identify the target population, being the units from which, or upon which survey data should be collected depending upon the nature and objectives of the survey, including individuals, households, schools, hospitals, facilities, and farms, among others. Moreover, to facilitate the process of identifying the population, the following questions shall be addressed:

What are the interests of the party requesting data? This question helps to identify the type of units to be studied in the survey and the definitions of the main features of these units.

Where do these units exist? This question helps to determine the geographic location of these units; the survey's objectives could be tackling a specific group of units in the population that could be located in a particular region or area of the country.

Is there a time reference for data? This question helps to determine the timing for implementing the data collection for the survey. It could be in a specific month, week or season.

## D. Analysis Plan

Once all variables to be studied and measured have been planned, the details required from each of these variables should be specified. In addition, the plan should include the required form for the results, whereby the required standards, totals, averages, rates, among others, are determined.

Moreover, the plan shall also include the survey output tables , enabling determining the best level of detail of data and, thus, determining the best objectives thereof.

The aforementioned may be included under the analysis plan, which would enable determining whether the findings cover gross or subtotals. Furthermore, such a plan is also used to design the presentation of the final tables of results based upon the structural or simple variables included therein. This analysis plan also facilitates creating the survey questionnaire.

Accordingly, it is necessary to clearly determine the survey objective during the survey design and elaboration stage to avoid facing obstacles during implementation. Hence, the following provides the most important inquiries which must be taken into account upon determining the survey objectives and the required data:

- What are the data required for the survey in general?
- Who is the data user, and how will such data be used?
- What are the terms and definitions which will be used in the survey?
- What are the detailed topics which the survey shall cover?
- Were any analysis plan and deliverables tables elaborated?
- What is the accuracy level of the required data?

## 2. Factors Impacting Survey Objectives

Several Factors influence the formulation of survey objectives, such as the quality of variables estimates, where it becomes necessary to identify the accuracy of the results. Accuracy of data is identified through the standard error of key variables in the survey, considering that a high level of accuracy requires a large sample size, which sometimes cannot be financed. Hence, a lower level of accuracy for key variables or lower levels of detail becomes inevitable ,that will affect on the survey objectives..

The most critical factors impacting the accuracy of the survey and should be reflected in terms of setting the survey objectives include the following:

- The size and design of the sample;
- The response rate;
- The time needed for carrying out the survey;
- The burden of obtaining the required data from the respondent.

## 3. Survey Planning and Managing

The procedure for survey planning and management is considered one of the most important factors that pave the way for a successful survey. Establishing an effective management structure leads to a clear understanding of the objectives that are to be achieved and a comprehensive view of the operational mechanism to achieve them. Planning and management are considered two main processes in ensuring the success of a survey and the achievement of its objectives. Through the planning process, requirements of the task and the human and financial resources are identified, and a timetable for the activities is elaborated. The aim of this part is to explain the procedures for developing the survey plan, highlighting the method to be followed in forming the survey team, as well as effective and good planning for the survey.

### 3.1. Survey Planning

Survey planning is a key part of any integrated statistical operation. Hence key regulations and specifications should be set in terms of planning the survey to ensure the success of the statistical process.

- The survey data and statistical information should be usable.
- The survey data and statistical information should be prone to evaluation regarding these six following properties: accuracy, importance, timing, accessibility and interpretability, and consistency.
- Justifications for respondent burden or the public's potential reaction should be made available.
- In designing the questionnaire, specific points should be taken into consideration, such as the time taken by the respondent to fill the questionnaire, the time needed for the survey implementation, the design of the questionnaire, and the method of data collection.
- Survey findings that are to be disseminated should reflect an objective statistical viewpoint. The reliability of these statistics shall be verified and confirmed.
- Data should be consistent with the correct classifications, standards, practices, and methods that are adopted in the field of statistics.
- Statistically correct and efficient methods shall be used. The correct practices and methods shall be abided by to avoid wasting resources. Work shall be carried out efficiently, and the professional reputation of the statistical entity shall be preserved.
- The survey should be carried out within the approved budget and the allocated time frame.

### *3.1.1. Stages of Survey Planning*

Survey planning is carried out in various stages to raise the level of accuracy and efficiency, starting with proposing the survey, identifying the main requirements and goals to be achieved by the survey, and the data to be collected. Moreover, efforts are exerted to ensure that the data are more accurate.

**A. Proposing the Survey**

This is the first step in planning for a new survey, whereby the required data to be obtained in terms of the survey is identified along with the deliverables to be produced. The proposal is developed following the preliminary study of the data available from administrative records or previous sample surveys and after consulting experts and specialists from all related fields to discuss the rationale for the study and the feasibility of its implementation.

**B. Developing the Feasibility and the Primary Plan for the Survey**

This is the main stage of planning for a new survey, where the overall actual cost of the survey is assessed compared to the estimated cost. This planning stage also aims at the following:

- Identifying the scope of operation, the goals, and the cost;
- Identifying and assessing the possibility of accessing the current data sources and determining the gaps and lacks in terms of existing information (current surveys and administrative data);
- Specifying the statistical frames and units for drawing samples and collecting data;
- Identifying the statistical methodology for the survey upon referring to similar previous sample surveys and carrying out benchmarking processes;
- Submitting a preliminary assessment of the cost, time frame, feasibility, and burden of response, including the evaluation of the appropriateness and impact of drawing samples on the estimated cost and quality requirements;
- Developing the feasibility and planning report while ensuring the coverage of all stages of the survey, including setting the goals and the framework options, designing the sample, collecting, processing, and disseminating data.

**C. Developing the Detailed and Alternative Plans**

In this stage, each team develops the detailed plan for its assigned tasks and coordinates with the other teams in the following:

- Developing plans for designing, implementing, and assessing the activity and time frames, and estimating the required resources and the cost of implementation for each team at every stage of the survey;
- Reviewing all the plans submitted by the teams, as well as identifying the inputs and deliverables for every team and every stage;
- Implementing the required developments as one of the main steps for the plans submitted by the other teams;
- Establishing links and consistency between the plans of the different, as well as the relevant inputs and deliverables;
- Developing the detailed plan of the survey and the detailed time frame;
- Developing the proposal for the pilot of the survey;
- Reviewing the budgets and modifying plans as requested.

In addition to the elaboration of the detailed plan, this stage also relates to the development of an alternative plan ( Plan B) tackling any happenstance during the survey implementation, given that the implementation of some survey stages may not be carried out according to the set plan, whereby, for instance, the response rate may be lower than expected, which may consequently impact the survey budget and completion plan, or the data quality may not meet the required standards.

### *3.1.2. Managing the Survey*

In line with the most common and efficient methods of survey planning and management, the project management principle is applied by determining the key roles, tasks, and deliverables to be achieved within a specific time frame and determining the relevant human and financial resources. Regardless of the method used in survey management, this process also encompasses organizing, guiding, monitoring, and controlling the survey.

**A. Survey Team**

The survey team consists of members enjoying the technical skills required for planning, managing, and implementing the survey and working on reviewing all proposals and budgets, and efficiently and effectively studying the needed options and strategies. Even though team members are assigned individual responsibilities, the team shares the responsibility of achieving the survey objectives within the set time frame. The team is often formed of the following members:

**Survey Manager**

The survey manager is responsible for managing the survey and ensuring the achievement of the survey objectives within a defined budget and a set time frame. He is expected to identify the human and financial resources, develop the implementation plan in coordination with the other team members, and monitor the appropriate usage of resources and the workflow.

**Statistical Specialist**

The person is responsible for the statistical content of the survey. In addition, his role includes ensuring the achievement of the objectives through using the survey data and is especially competent in terms of collecting and analyzing the relevant historical data (for planning and development), developing the content and testing the questionnaire, preparing and designing the statistical deliverables, developing and analyzing data, and developing the analytical text.

**Methodologies Specialist**

The person in charge of designing and developing a statistical methodology to be used in the survey, designing quality control methods, and the mechanisms for data auditing. He is also responsible for ensuring efficient and appropriate compliance with the statistical categories, concepts, and definitions.

**IT Analyst**

The person in charge of designing and developing the computer systems and software, modifying the current software required for implementing the survey, ensuring that these systems are aligned with the specifications provided by the other team members and participants, providing technical consultations for all other team members on any computer-related matters.

**Data Collection Specialist**

The person responsible for managing the data collection process according to the adopted methods and approaches within the set time frame and under the survey quality standards, as well as for planning and coordinating the recruitment, employment, and training processes, and identifying the logistic support required to implement the survey.

**Other Members**

Larger projects, such as the Population Census Project, might require more members and coordinators. In terms of projects, a communication and advertisement manager, a logistic support manager, specialist members might be appointed, and technical and coordination committees might be formed to ensure the workflow as per the goals and the time plan.

# 3 | Questionnaire Design

The statistical Questionnaire is the tool used in a wide range of studies to collect information and data relating to statistical variables for a specific issue or phenomenon. The data and information on variables being collected could be economic, social, and demographic, among others. The questionnaire consists of several questions tackling several variables that are produced in the form of a hard or soft copy, whereby such questions are addressed, sent, or handed to the selected respondents under the survey sample.

## 1. Questionnaire Design

This stage includes several steps, being the following:

- Identifying the type of the required information: Designing the Questionnaire is carried out in line with the general framework of the survey objectives, whereby such framework encompasses the main theme and sub themes of the surveys, are rationally classified, and list, under each of them, the questions relevant to them. In general, questions vary according to the required information, whereby the form may include questions that relate to facts, such as age and academic qualification, other questions on opinions or points of view, or questions about general information.
- Determining the method of data collection: The length of the questionnaire and the way questions are introduced depending on the data collection methodology. For instance, the surveys that adopt the self-completion data method usually have less complex, shorter forms with more explicit instructions and questions than those used in personal interviews.
- Identifying the properties of the sample units: When formulating the questions, properties of the sample units should be considered because these could impact the terms or the complexity of the language used in terms of questions. Hence, questions prepared for the public should be easily comprehensible to all participants, but if the survey is targeting professionals, the relevant professional or technical terminologies could be used.
- Formulating the questions, identifying their content along with the objective of each of question, and the time and effort needed to answer such question is necessary, whereby the number of questions shall be kept at a minimum to lessen the response burden.
- Designing the survey output tables to develop a clear perception of how to deal with the data to be obtained, whereby such a step is carried out before creating the survey questionnaire.
- Determining the type of the questions is necessary for being open-ended, multiple-choice, or a mix of both questions.
- Identifying or developing a detailed coding system upon formulating the questions.
- Each questionnaire shall bear a unique identification number which facilitates referring to the original questionnaire to verify that the information in the sampling data file matches the information listed under the form.
- Reviewing the previous questionnaires: Questionnaires from previous surveys form a good source of information in terms of designing a new questionnaire. Testing the questions used in similar surveys is a useful starting point for designing the questions.
- Consultations with data users: The main objective of data collection is to deliver data that meets the needs of various data users. It is therefore important to communicate with data users and consult with them on formulating the survey objectives, whereby such step is carried out as follows:
    - Identifying the key users and calling for meetings during the stage of developing the survey questionnaire in order to hold discussions with them, identify their needs, and reflect them under the form;
    - Developing documentation records for all the inquiries and requests made by data users;
    - Designing a brief feedback form to serve as an attachment to the proposed survey questionnaire to obtain feedback on the quality of the product and its capacity to meet users' needs.
- Developing the questionnaire in its final format: Under this step, the format and layout of the questionnaire are processed per the general guidelines adopted in this respect. These guidelines will be listed in detail under the Form Elaboration General Guidelines section.

## 2. Preparing and Reviewing the Questionnaire

### 2.1. Reviewing the Questionnaire

It is essential to review the questionnaire because this is a key procedure for testing and assessing it. All possible errors and problems are identified during the review procedure, such as spelling and grammatical mistakes and inappropriate forms of questions. It would be instrumental at this point to assign the revision of the document to individuals who are not directly participating in the survey, as well as a group of experts in terms of research and study methodologies, a group of individuals experienced in developing questionnaires, and individuals from the survey population.

### 2.2. Performing a Pilot Test of the Questionnaire

Testing the questionnaire is another necessary procedure to help identify and detect any errors or biases. Both the enumerator and the respondent might face problems or issues arising from the content of the questions, such as distortion and vagueness of the overall meaning of some questions, or the misinterpretation of some expressions and concepts, or problems related to moving from one question to the other, thing that might lead to losing some data.

There are several methods for testing the questionnaire based upon small samples of the targeted population. These methods are discussed in the following sections:

#### A. Pre-Data Collection Pilot Testing

During this testing stage, several procedures are applied before the data collection process, and that during the pilot stage of elaborating the questionnaire, such procedures involve teams and workgroups and preliminary interviews.

#### Team and Workgroups

A workgroup is a small number of participants carefully selected from the target population and data users, who are to discuss the survey topic that has been selected. Such discussions will provide the survey designer with useful information about how these individuals relate to the topic and an insight into their thoughts and life in terms of such topics. Such detailed information cannot be known from the statistical survey data. Hence, such insights are included in the questionnaire and reflected therein.

#### Preliminary Interview

These are personal interviews with a small sample of respondents, in which the preliminary version of the questionnaire is used; respondents are to express their opinions through their answers to the questions. This method is very useful in selecting questions in verifying the following:

- The way the respondents think as well identifying the reasons behind the non-response to some questions;
- Helps to assess the validity of the questions and formulate clearer constructions through follow-up questions raised by the enumerator. For example, he might ask how the respondents selected their answers or how they interpreted the response options. He might also ask about the meaning of a statement, whereby the respondent reformulates the question in his own words, to know whether the respondent has fully understood the question and interpreted it as intended; Assessing the design of the questionnaire from the respondent's viewpoint.

#### B. Pilot Testing During the Data Collection Process

Pilot testing (prior testing) is an opportunity to test the survey design based on a small sample with matching properties to those of the survey sample, during which all stages of the survey are to be implemented. This process is helpful in testing the questionnaire in terms of the following:

- Identifying the rate of response to the questionnaire;
- Assisting at the identification of ambiguous questions;
- Identifying common errors in terms of formulating the questions and the question order issues that could lead to bias;
- Testing the effectiveness of the various components of the questionnaire: introductions, guidelines, section titles, etc.;
- Detecting aspects of bias in response, if any.

### 3.2. Classification of Questions Types

Questions listed under the statistical questionnaire may vary according to their form, the type of answers required, and the type of variables it includes, which includes quantitative or qualitative variables and in terms of open-ended answers or multiple-choice questions.

#### a. Classification Questions as per their Form

One way of classifying questions is under their categories and types and thus, as per their answers. Hence, we can distinguish two types of questions: open-ended questions and multiple-choice questions.

**Open-Ended Questions:** These are questions directed to the respondent, leaving them the freedom to answer in their own words, thus according to their beliefs and trends, whereby they are granted the freedom of expression to give their opinion.

**Multiple-Choice Questions:** These are questions directed to the respondent and require selecting only one of a specific number of predefined answers set by the survey designer. In some cases, the objectives of the question would require the selection of more than one possible answer to it.

Each of these question types has its advantages and technical requirements. In comparison, open-ended questions offer the respondent the opportunity to provide adequate and detailed answers to the questions; such typefaces challenges in coding answers so that they are easily entered into the computer for tabulation and analysis. However, multiple-choice questions are limited to specific answers leaving no space for the respondent to explain details, but answers are easily obtained and pre-coded.

On the other hand, multiple-choice questions could be classified into:

**1. Nominal Questions:** Used whenever the answer to them relates to types/classes, given such type/class relates to one category only, such as: What is your gender?. Or more than one category.

| | |
|---|---|
| Male | Female |

**2. Measurement Questions:** Used whenever there is an attempt to identify the stance, opinion, or feeling of respondents in terms of a specific matter, such as: In your opinion, how important is the high school grade in terms of the student's excellence in university?

| | | | | |
|---|---|---|---|---|
| Very unimportant | Unimportant | Neutral | Important | Very important |

**3. Ordinal Questions:** Used whenever their answers are a value of ordered variables which are ranked according to specific factors or properties, such as: Kindly arrange the following characteristics of a team leader from (1) to (5), based on importance, where (1) stands for the least important and (5) represents the highest level of significance.

| **Characteristic** | **Rank** |
|---|---|
| Honest team leader | |
| Team leader who obtains the necessary resources for the team | |
| Team leader who calls for team spirit | |
| A strong and firm team leader | |
| Team leader who motivates the team members in an exemplary manner | |

**4. Numerical Questions:** Used whenever their answers are values related to numerical variables, such as age, income, and expenses, such as: How old are you in years? What is your average income in Dirhams?

**B. Classification Questions as per their Content**

The questions could be classified or arranged by the type of information it includes, as follows:

1. Demographic Questions: Used whenever their answers relate to demographic variables, such as age, gender, profession, and academic year, among other descriptive information.

2. Behavioral Questions: Used whenever their answers include information on practices and behavior, namely how a person behaves, such as: How many hours daily do you spend watching TV?

| | | | | |
|---|---|---|---|---|
| I do not watch TV | One hour | Two hours | Three hours | Four hours or more |

3. Opinions Questions: Used whenever their answers explain how a person would think under a specific circumstance or situation, such as: Are you always happy? Kindly circle your answer.

| | | | | |
|---|---|---|---|---|
| Very happy | Happy | Neutral | Unhappy | Very unhappy |

### 3.3. Guidelines on Formulating Questions

Designing the questionnaire and formulating questions should be done to ensure the clarity of questions to the respondents. It is essential that the respondent understands and comprehends the details of the questions as intended by the questionnaire designer to avoid any negative impact on the quality of the data. Hence, the following provides some general guidelines on formulating questions:

- The language of the questions should be simple and appropriate for all levels of respondents.
- Questions shall be formulated in a way that does not lean towards a specific answer.
- The structure and wording of the question should not be subject to interpretation or the possibility of understanding it in multiple ways allowing for more than one answer to the same question.
- Dual questions (introducing two ideas within the same question) shall be avoided, such as; Are you planning to leave your car at home and commute to work by bus next year?
- All appropriate possible answers shall be provided in terms of multiple-choice questions.
- Some questions shall be listed in more than one form to make sure that the answers are accurate.
- The questions should not require deep thinking or complex calculations.
- Ambiguous terms shall be explained.
- Codes and abbreviations shall be defined. For instance, (LLC) would be defined as a (Limited Liability Company) abbreviation.
- The use of open-ended questions and questions that might cause specific reactions among respondents shall be limited, and they will fall at the end of the questionnaire .

### 3.4. General Guidelines on the Questionnaire Preparation

This stage is concerned with the questionnaire's layout and producing it in an appealing way. The following points should be considered when preparing the questionnaire:

- The questionnaire should be designed simply and efficiently so that the enumerator can obtain the data quickly and simply that does not produce boredom on the respondent's side. This will help to improve the accuracy of the data obtained.
- Clear linguistic terms which are familiar to enumerators and respondents should be used within the statistical questionnaire.

- The questionnaire should neither be extended nor include more questions than those needed to reflect the survey objectives. More time is required to provide data for lengthy questionnaire, which may reduce the accuracy and efficiency of the data.
- The introduction to the questionnaire should include the following: the survey's title, its objective and importance, and finally, identify the entity carrying out the survey. It shall also reiterate data confidentiality and guarantee that data will solely be used for statistical purposes.
- The transition between questions should be easy and clear, leading to no confusion on the respondent's side upon completing the questionnaire.
- Instructions on transitioning between questions shall be provided to facilitate the process of completion for the interviewer.
- The guidelines provided to both respondents and interviewers shall be brief, clear, and accessible, and the definition of the terms shall be provided either at the beginning of the questionnaire or alongside the related questions.
- The questionnaire shall be divided into consistent groups of questions, which shall ear bear a title.
- Questions should be arranged and numbered in a way that facilitates following up on both updating the data and electronically entering data into the computer.
- Upon designing the questionnaire, a page or a margin is to be allocated for the comments written down by the enumerator.

# 4 | Sampling Design

## 1. Introduction

Identifying the appropriate sampling technique is considered one of the key stages for implementing the sampling survey. A sample, as commonly understood, is a subset of the population that has been designed and selected according to statistical principles, which would ensure that each unit of the population has an opportunity to appear in the sample. Moreover, it is known that if all units of the population have an equal chance of selecting in the sample, the latter would then be called a "simple random sample," which is the most straightforward and easiest of all statistical sample designs. In contrast, other designs are more complex, such as systematic samples, samples proportional to the size of the population, stratified samples, cluster samples, and multi-stage samples; all of which are instances of probability sampling which gives every unit of the population a chance or probability to appear in the sample. On the other hand, non-probability sampling is used for convenience, does not necessarily offer each population unit a chance to appear, and requires no sampling frame for selection.

This chapter discusses statistical sampling techniques, as well as the absolute and relative statistical sampling errors, as measured by the coefficient of variation expressed as a percentage CV%, which are calculated for evaluating the efficiency of the results produced by the sample. More comprehensive details are available in the guide of statistical sampling produced by the Statistics Centre – Abu Dhabi, which is available on Statistical Methodologies Portal, in SCAD Website

## 2. Key Concepts and Definitions

This chapter tackles the key concepts and definitions related to both the theoretical and practical sides of sample design and sample selection, which are consistent with international terminology and concepts in this field:

- **Population:** is the collection of all statistical units from which the survey sample will be drawn. These units should be clearly defined so that they share one property or a set of common properties. Most populations consist of statistical units that change over time (renewable populations), although some others are static populations that do not change over time.

- **Sampling Technique:** is a technique used to select a sample from the population and process them on the statistical level so that the results from the analysis of the sample will represent the statistics of the population that are to be estimated.

- **Random Selection:** The process of selecting units from the population in a way that excludes any personal control in selection process, to ensure that an opportunity is granted to all units in the population to appear in the selected sample.

- **Frame:** A list or file that includes all units of the statistical population that will consist of contact details, often including the names and addresses of statistical units and some related information. In other words,the frame is the set of documents that helps us reach the statistical units to collect data from them.

**Sample:** A subset of the population that is selected using statistical sampling techniques to represent the population that is being studied. For the sample to be representative of the population, it must include the properties of the population in a way that enables us to generalize its results to estimate the key properties of the population.

**Sample Design:** is the process of selecting the appropriate method from all the various types of sampling methods to reach the sample to achieve the targeted results.

- **Sampling Proportional to Size:** is a sampling technique where the probability of selecting each sample unit is proportional to the size of that unit concerning the property under study. For example, the size of an establishment is measured by its number of workers. Upon selecting a sample of establishments through the method of sampling proportional to size, a higher probability or opportunity is granted for the appearance of larger establishments, being those with a larger number of workers.

- **Sampling Unit:** is the basic unit of the population that will be selected for the sample. In other words, it is the unit from which required statistical data or information is collected.

- **Primary Sampling Unit:** the sampling units that are drawn in the first stage of a multi-stage sampling design. Usually, a primary sampling unit would represent a cluster of units, thus being a primary sampling unit including a group of secondary sampling units.

- **Secondary Sampling Unit:** the sampling units that are drawn in the second stage of a multi-stage sample design; each secondary sampling unit is considered a part of its primary sampling unit.

- **Analyzing Unit:** the unit used for analyzing the statistical data collected to achieve the survey objectives. It is possible that the analyzing unit is the same as the sampling unit used in sampling or may be different.

- **Non-Coverage Errors:** The errors that may arise during the elaboration of the sampling frame, where they are related to the lack of the frame or over-inclusiveness of the frame. Problems may occur because of introducing individuals to the frame that do not belong to the population or failure to describe the individuals within the frame appropriately. Non-coverage errors are divided into two types: first, under-coverage, or failure to include individuals that should have been included, and second is over-coverage, which means including individuals that should not have been covered.

- **Non-Response Errors:** is the failure of some respondents to answer the complete questionnaire, whereby such non-response can result from rejection or other reasons. This type of non-response is called "unit non-response," whereas the "partial non-response" occurs when the respondent only declines to provide answers to specific questions.

- **Standard Error:** This is the square root of the estimated sample variance divided by the sample size. Sample variance is the average of the squared deviations between the values of sample units and the variable's mean value for these units.

- **Relative Standard Error:** is the standard error divided by the mean of the data for which the standard error has been calculated and is equal to the coefficient of variation.

- **Optimum Allocation:** is one of the techniques for allocating the units of stratified samples into the various strata so that the sample size in each stratum is directly proportional to the size of the stratum,the variation within it, and the cost of collecting the sample unit in the stratum.

- **Nyman Allocation:** is one of the techniques in stratified sampling for allocating the sample across the strata, where the sample size in each stratum is directly proportional to the size of the stratum and the amount of variation within the strata.

- **Proportional Allocation:** is one of the stratified sampling techniques for distributing the sampling units amongst the various strata, where the sample size in each stratum is directly proportional to the population size of this stratum.

- **Bound of Error:** is the standard error value multiplied by the Z-score value or t-student value , within certain confidence limits.

- **Weighting:** is the calculation of raising factors based on probability of selecting the sampling units, aimed at obtaining the values of the parameters estimated for a population upon using the sampling survey results.

## 3. Types of Samples

In terms of sample selection ,statistical samples are divided into two main types:

### 3.1. Probability Samples

These samples are drawn in accordance with the laws of probability, where the population units are drawn successively under a known probability, by one of the probability Selection methods. In probability sample design the sample outputs can be raised or generalized to all units of the population through calculating the sampling weight of each sample unit, whereby the weight of the sample is the inverse value of the probability of the unit being drawn in the sample from the population.

There are many types of probability samples that depend on selecting the most appropriate type thereof as per the nature of the population in terms of consistency or variance and the type of survey or the variables that are being subject to research. The following provides a detailed overview of each type of sample:

**A. Simple Random Sample:** It is the base of the probability sample and relates to all the other types of samples. In summary, it is defined as granting each units of a population an equal chance of selection in the sample. Usually, samplers use the computer software of random cases lists for selecting a random sample. As previously mentioned, random sampling means that each unit of the population has a known and equal chance of representation under the sample, whereby such probability is known as n/N, where n refers to the sample size and N refers to the survey's population size.

For instance, if the population size is formed of 60,500 individuals and a sample of 1,800 individuals is to be selected, the probability of representation of each sampling unit of the population is reached by dividing 1,800 by 60,500, which is equal to 0.02975.

**B. Systematic Sample:** A systematic sample relates to systematically selecting sampling upon randomly selecting the first sampling unit, calculating the systematic period through dividing the population size by the sample size, whereby the systematic period is cumulatively added upon drawing each sampling unit, and the serial number is assigned to the individual holding the cumulative number.

For instance, in the selection of a sample of 100 individuals out of a population formed of 10,000 individuals, the systematic period would be calculated by dividing 10,000 by 100, equal to 100. Upon assuming that we have randomly selected the number 14 as a starting point, the sample would be formed of the individuals holding the following numbers: 14 / 114 / 214 / 314 / 414 and so forth until 100 sampling unit is reached.

**C. Stratified Sample:** A stratified sample is used in terms of ensuring that various groups of the population are represented under the sample. The main idea of a stratified sample is that the information on the survey's population is applied to divide such populations into groups with common properties. For instance, let us assume that the population was formed of 700 individuals of nationality A, 200 individuals of nationality B, and 100 individuals of nationality C. In case a random sample was drawn and formed of 10 individuals, a stratified sample formed of 70 individuals of nationality A, 20 individuals of nationality B, and ten individuals of nationality C, would represent the group in a better way.

It is important to note that it is crucial to be cautious in dividing the population into many strata, as such action would increase the sample size. The stratified sample may be proportional upon selecting a fixed number from each group, whereby the sample size of each stratum would be proportional to the population size under this stratum. If the overall number of population individuals is different in each stratum, the stratified sample would be considered disproportionate.

**D. Cluster Sample:** A cluster sample is used in higher-level studies as it contribute in reducing the survey costs during field work execution.. A cluster sample relates to selecting large groups known as clusters, whereby the clusters as a primary sampling units are selected in the first stage with probability proportional to the size, in the second stage the secondary sampling units are selected from each cluster by simple random sample method or systematic method.

### 3.2. Non-Probability Samples

Non-probability sampling is defined as a sampling technique in which the sample is selected based on the subjective judgment of the sampler rather than random selection.

In general, samplers usually prefer to select samples according to random or probability sampling rather than to non-probability sampling, whereby the former is considered more accurate. However, the implementation of probability sampling may not be helpful, on a practical or theoretical level, in terms of applied social research. In these cases, we may apply non-probability sampling techniques, which include the following:

**A. Convenience Sampling:** This type includes different ways of sample selection, such as selecting individuals who are coincidently met on the street. It is a method adopted by TV channels for obtaining an overview of public opinion. Moreover, under any circumstances, sampling units are drawn from volunteer groups. However, the inconvenience of such sampling techniques is that there is no evidence proving that it represents a population.

**B. Purposive Sampling:** The sample is selected with a specific purpose, whereby we usually have a group to research, such as students or employees, among others. A purposive sample is useful in cases where we would like to draw the sample quickly. It also helps in identifying the opinions of the target population. However, under this sample, it is possible to grant a bigger weight for the easier groups to reach the study's population. Moreover, all the following sampling techniques may be used as sub-types of the purposive sampling technique.

**C. Quota Sampling:** the sample is drawn in a non-random manner, thus according to a specific quota. There are two types of sampling in this respect: the proportional and non-proportional type, whereby, under a proportional sample, we would want to represent the key characteristics of a population through a sample appropriate to each of the relevant properties. For instance, in case we know that a population is formed of 40% of women and 60% of men and that the sample we need would be formed of 100 individuals, we would keep on drawing sampling units until such rates are reached. However, in case 40 women are included under the sample, and 60 men were not, we would keep on drawing sampling units from men only, while selecting women would be prohibited, even in the case where women would meet the selection terms and conditions. On the other hand, non-proportional sampling is less complicated as it allows setting a minimum of sampling units for each category regardless of the proportionality between the sample size and the population properties.

**D. Snowball Sampling:** In this technique, the first step would be selecting an individual who meets the criteria set for the sample selection, whereby the individual would be asked to select other individuals sharing the same criteria. Although this technique does not accurately represent a population, it is useful in some instances where reaching individuals from the survey population is difficult.

# 5 | Statistical Data Collection

The data collection stage is considered one of the key stages of statistical work because of the consequent quality of the indicators obtained. It accounts for the highest cost of all the survey stages, This chapter will discuss the processes and procedures performed during the data collection process.

## 1. Studying the Survey Plan and Objectives

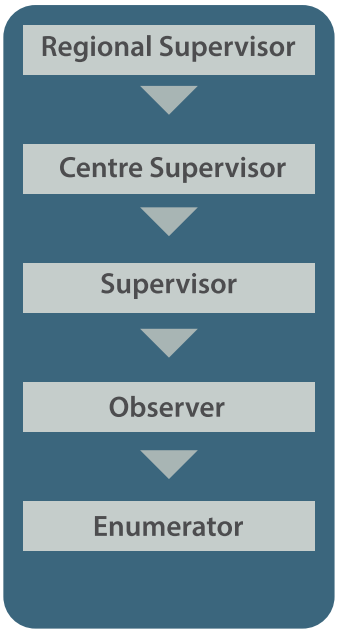
In this stage, both the plans and objectives of the survey to be implemented are checked, the questionnaire, and the time frame for the survey, are reviewed. During this process, the following points are identified:

- Determining the nature of the survey (economic, household, environmental);
- Identifying the appropriate times to undertake fieldwork;
- Identifying the estimates of the numbers of required enumerators and their qualifications. For instance, economic surveys need enumerators competent in economics, accounting, management, or other similar specialties;
- Identifying the logistic needs (specifying the supervision centers and their locations and identifying any requirements in terms of transportation, among others).

Risks in this stage include the possibility of incorrect estimates for the requirements of the survey, which might consequently lead to budget shortfalls and project failure, or the identification of alternative solutions that might influence the quality of data.

## 2. Identifying the Requirements of Human Resources and the Organizational Structure

Based on the preliminary study of the survey and the sample size that has been reviewed and studied, the labor force resource for the survey is determined. Then, the resource is distributed based on the sample and the geographic regions, considering the survey time frame. Accordingly, the number of enumerators is identified. Hence, the technical and administrative resources are identified, usually a ratio of 1:4, which means that one supervisor for every four enumerators, whereby the supervisor is to supervise them on the technical and administrative levels. Afterward, supervisor resources, regional supervisors, and centers supervisors are identified based on the sample size, , and geographic coverage of the survey.




It is worth noting that technical supervision is carried out at levels that are lower than the senior administrative levels. In other words, whenever administrative work increases, technical work decreases, whereby technical supervision is at its peak within the functions of both supervisor and observer. It is also worth mentioning here that the administrative hierarchy could be modified, and in some cases, some administrative positions could be ignored according to the survey size and the number of personnel working in it, and its time frame.

### 2.1. Selecting and Assigning Data Collection Team

The stage of selecting and assigning data collection team goes through the following main stages:

**Personal interviews:**

Upon approving data collection, personal interviews are carried out for personnel nominated to work as interviewers. There are several quality standards to be followed when selecting and training the temporary resources assigned to implement the data collection stages, whereby these standards vary according to the type of survey to be implemented. As such, standards of demographic surveys are different from those of economic surveys, but in general, the following criteria should be available in all personnel working in the survey:

1. Academic qualifications matching with the type of the survey;
2. The overall image and the proper appearance to best represent Statistics Centre – Abu Dhabi's image;
3. The practical experience within the same field; and
4. Compliance with the instructions and considering the administrative structure of the project.

**Training:**

The training stage is considered one of the most essential stages in selecting the interviewers, whereby interviewers are trained on the technical aspects related to methodologies and questions included in the questionnaire and the various concepts and definitions. They are also offered training on the legal aspects and the ethics of dealing with the sampling units. The following is illustrating the aspects of training:

- Training on the methodology of the survey;
- Training on the concepts and terms used in the questionnaire;
- Technical training (using tablets and other devices); and
- Training on the ethics of dealings with population units, the rules and regulations of fieldwork, and the relevant legal aspects.

**2.2. Identifying Technical and Logistic Requirements**

The process of identifying the technical and logistic requirements are related to the following:

- The nature and type of the survey;
- The number of personnel working in the survey; and
- The number of proposed supervision centers.
- In addition to the aforementioned, logistic needs are divided into two parts:
- General services, including all the items pertaining to the readiness of the supervision centers and work tools; and
- Information Technology, including all the items pertaining to computer devices, networks, printers, scanners, software, and devices used in terms of the electronic data collection.

**2.3. Establishing Back Offices and Control Systems**

After identifying the organizational structure of the survey, a system for back offices is to be developed to ensure the flow of data collection and reflect the daily accomplishment in the field so that it enables data collection team to carry out their assignments, including data entry, auditing, and approving questionnaires, and as it also enables those in charge of the survey to follow-up on the progress of the survey, as well as the flow of work through providing them with daily performance reports. One of the most important conditions to be met by the daily report is simplicity and the ability to reflect the actual image of reality in the field, whereby such reports would organize the workflow.

Before starting data collection stage, it is necessary to examine all tools used in the survey, such as the electronic questionnaires that are downloaded to IPADs, input software for surveys, in addition to the systems of the back office. In terms of this process, fake data is entered for testing the flexibility of the program and ensuring the implementation of rules of consistency, and databases are tested, as well as their accuracy and alignment with the requirements.

Accurate examination of systems and tools consequently leads to avoiding the technical problems which the teams might face upon implementing the process of data collection.

**2.4. Data Collection Plan Elaboration**

During this process, a plan for data collection is to be developed, whereby such a plan would become the drive for all activities related to data collection. The plan includes collecting all data and information in the survey and the targets about daily progress and achievements, and the dates of data collection. Once interviewers and supervisors are selected and assigned to their various areas:

The sample units lists are handed to each supervisor/observer, whereby the sample shall list the names of the pools, streets, buildings, and locations of the units to be interviewed clearly, including all the relevant details to

facilitate access to such sampling units for the supervisor/observer who would be able to deliver such data to his team before beginning the data collection process.

The forms of daily progress reports shall be handed.

A quick review of the survey methodology and data collection mechanisms shall be performed.

## 3. Data Collection Stage

During the stage of data collection, the interviwer is to communicate with the respondents, whether individuals or establishments, among others, and arrange for holding a meeting with them, then meet with them to manage the process of filling the data in the questionnaire and ensuring the entry of all required data.

The field supervisor is to regularly follow up on the work in the field with the personnel through the following:

- Performing field check and verification to ensure their presence in their appropriate work spots;
- Receiving daily reports to follow-up on their field performance;
- Comparing the daily achievement with the predefined expected achievements;
- Looking into the problems related to the daily performance and meeting with the team to clarify the weaknesses to identify means for addressing them;
- Reporting on the problems facing the work area to the senior management to identify their failure and attempt to address them;
- Dealing with non-response (whole or partial);
- Ensuring the coverage of all sample units before closing the data collection stage; and
- Auditing the interviwers' work to exclude the presence of repetitive errors, whereby such errors, if any, are notified to the senior management for addressing them.
- In addition to the aforementioned, the supervisor is to monitor the following tasks by the end of the survey:
- Ensuring the coverage of all sampling units prior to announcing the end of the time allocated for data collection; and
- Reporting on the sampling units that have been visited and the special cases, for example, non-inference cases, non-response cases, closed cases, and refusal cases.

## 4. Managing the data collection quality

The success of the data collection is measured by the level of data quality and the accuracy of the procedures, as well as the results it produces. It should be noted that the concept of quality of statistical data is a wide and comprehensive one that relies on the accuracy and efficiency of statistical data. This concept has developed to become more comprehensive for the world's international institutions and national statistical centers.

### 4.1. Quality During the Data collection Stage

The data collection stage includes several technical procedures that the quality control team should focus on to ensure the appropriate implementation, some of which are related to the behavioral and administrative aspects and some others to the technical procedures.

**Stages of applying quality control in terms of the administrative and behavioral aspects include the following:**

- Performance Monitoring: The quality control team retrieves a random sample from the sampling units and re interview such sample with the interviewer and supervisor to assess their method in collecting data and their ability to apply the technical instructions given to them.

- Teams' Behavior and Performance Monitoring: The quality control team assesses the data collection team upon measuring the extent of their cooperation, compliance, behavior, and appearance. In this case, the supervisor evaluates the observer and assesses the interviewer. Similarly, the interviewer assesses the observer, and the latter assesses the supervisor.

**Stages of applying quality control in terms of the technical procedures include the following:**

- Quality control throughout interviewer's work: It is the role of the quality control team to verify that the field enumerator is appropriately implementing the technical procedures to ensure that the data included in the questionnaire is correct and that all terms of the questionnaire have been covered. The quality control team is to take a sample of the questionnaires immediately filled by the interviewer to verify the accuracy and efficiency of the implementation method. This includes ensuring appropriate data entry included in the questionnaire to the data system.

- Quality control for coverage and comprehensive cases: (closure, non-inference, out of survey scope). It is known that the interviewer and the observer, should verify these cases, , whereby the role of the quality control team here would be to take a sample of these cases for their verification at the field and the office levels, to guarantee a high level of accuracy.

- Verifying the selection of the correct sample targeted for the survey: The quality control team takes a selected sample from the samples provided by the interviewer and the supervisor, to ensure that the interviewer has met the relevant requirement.

- Quality control during desk auditing: The quality control team is to focus on ensuring the appropriate implementation of technical auditing procedures adopted by the methodologies teams by taking a sample of the audited questionnaires.

**Ideal instructions and procedures to apply data collection quality control include the following:**

- Re interview the sampling units for check processes : It indicates on the accuracy and credibility level of the collected data during the collection stage, whereby a sampling unit is retrieved from the survey sample. The quality control team carries out this process to ensure the appropriate implementation of the recommended procedures and evaluate the accuracy of the interview. The respondent gives his opinion on the interviewer's behavior in dealing with him his ability to deliver the information.

- Technical circulars: It is essential to ensure that the technical circulars and errors and follow-up reports are implemented and delivered to the individuals who have proven to commit these errors

- Comparing the collected data from the current survey with that collected from other previous surveys to make sure the data is consistent , this processes indicates on the accuracy level of the collected data

- Holding periodic workshops and courses for both observers and supervisors to highlight the importance of the data collected by their teams, whereas such a step is considered a vital awareness-raising topic. These teams should have access to the results of the data they collect, their consequent indicators and the methods of data processing, and the emanating risks from obtaining inaccurate data and their reflection on the deviation of the values of the indicators.

- Preparing a statistical estimate for refusal and non-response cases of all types through establishing tables and figures used to classify these cases according to the location and the reason behind the refusal, according to the enumerator who collect the data. This assessment aims to link the reasons for non-response to the interviewer and the observer to provide recommendations on correcting the path of interviewer and addressing their weaknesses.

- Preparing weekly achievement reports targeting the measurement of the level of achievement based on the location and the sampling units type, according to the supervisor, the observer, and the interviewer

# 6 | Data Processing

## 1. Introduction

Data processing is all about processing the data obtained through the process of data collection in the surveys to make such data suitable to perform statistical operations later on, including tabulation and analysis of these data.

Ordering the activities related to data processing relies on the nature of data and the means of collecting it, the survey objectives, and the required level of data quality, among others. The following provides an example of the activities of processing data included in the questionnaire:

- Verifying the questionnaires to ensure that all the data that has been obtained was recorded in a correct and accepted way and that all comments made by the interviewer have been read and all the preliminary amendments required to verify the errors and contradictions within the questionnaire have been performed;
- Performing the necessary coding procedure for variables that need such process;
- Auditing the data after it has been encoded to make sure it is correct;
- Performing a detailed auditing process: In this stage, questionnaires containing errors are put aside for more investigation and verification of data, either for following up with the respondent or to compensate for the missing data;
- Detecting outlier values to identify the unacceptable values or those that fall out of the scope of correct answers; and
- Saving and storing data in the database to facilitate data processing during post-processing activities.
- Data processing is often performed within the framework of the aforementioned activities, such as data auditing and coding through electronic systems and software, which include automated auditing and coding rules.

## 2. Data Coding

The coding process means assigning codes or specific numbers to data and exchanging text answers (sentences or words) with codes bearing specific meanings according to statistical classifications and predesigned system that assigns a short code to the case. The coding process, being the first step of the mechanical processing of data, is usually performed under the supervision of a specialized team that uses various coding systems.

### 2.1. Self-Coding

Self-coding of closed questions means that each of the answers to the question has to have a label "code" that could be next to the answer on the form, also a small box or square is to be typed next to each question in the form so that the appropriate answer could be written inside it. This improves the efficiency and the speed of data entry to a great extent because the person in charge of this process usually focuses on the squares with the numbers inside them and the special codes for entry.

### 2.2. Guide Coding for Open-Ended Questions

Upon finalizing the data collection process, all open-ended questions are listed under one list, whereby the following steps shall be followed in terms of the coding process:

- The encoder shall elaborate an independent list for each question separately.
- All the answers for each open question shall be listed in alphabetical order.
- The encoder shall read and interpret the answers, whereby all similar answers shall be listed under one group to which a number or code is assigned.
- Then, the encoder shall register the number or code of each answer under the list and compare them with the list of answers.

### 2.3. Automated Coding of Open-Ended Questions

Due to technological advancements, the automated coding of open-ended questions has become possible. In general, this procedure resorts to using two files; one includes the answers to the open-ended question that is to be coded, while the other, being the reference file, contains the answers to the question and the code for every answer, based on a predetermined statistical classifications

The automated coding of open-ended questions is performed according to the following steps:

1. Analyzing the statement (the sentence): through dividing it into sub statements and key terms to help identify similar statements and words featured under the reference file.
2. Any unusual characters, such as numbers or special characters and repeated words, are to be omitted during the statement analysis process.
3. Then, the reference file will be subject to a search process to identify the key terms that were divided for comparison purposes. In case a word is found out to be identical to such terms, the code for this statement or sentence is to be copied and registered in the file, including the data of the study, yet if no identical word is found in the reference file, then we must search for a term which bears the closest meaning. In case no similar word has been found, the status "non-existent code" shall be assigned for it to be encoded at a later stage by a specialist team.

## 3. Data Entry

The process of data entry helps enter the answers into a machine-readable mode. There are two means of data entry:

### 3.1. Traditional Data Entry

This process uses paper forms to carry out interviews with the targeted sample individuals; data entry is to take place after completing the process of data collection (from the field) and reviewing, verifying, and coding it. The data entry team is to list the answers under electronic questionnaire that was specifically designed for this purpose, whereby the software designer shall consider the following in this respect:

1. The software should include all questions on the form and shall not miss any questions.
2. Entry screens shall be similar to the questionnaire structure as much as possible to facilitate the process for the data entry clerk.
3. It is essential to provide data auditing software upon finalizing the data entry process for all questionnaires, whereby this process is known as Batch Editing, the data auditing process after data entry, to ensure the accuracy and quality of the data that has been entered.

### 3.2. Contemporary Data Entry

This method aims at replacing the guide storage spaces with electronic storage spaces for storing the questionnaires and enabling their printing at any time.

***- Intelligent Character Recognition (ICR)***

This method is used upon applying the following steps:

- Designing a particular questionnaire using scanners so that it matches the process of distinguishing and reading data;
- Scanning all questionnaires collected in the field using scanners characterized by high speed and extreme accuracy is done by transforming paper forms into electronic copies, whereby optical scanning is carried out for scanning a set of files, called batches, to speed up the scanning process.
- The questionnaire is identified through its unique barcode. Then, the copies of the questionnaires are archived and stored.

*- Handheld Devices (Tablets)*

In this stage, an electronic questionnaire is to be designed for the handheld device instead of the traditional paper questionnaire, whereby such step has been better than the paper questionnaire in terms of many aspects, including the following:

- The quality of data collected through electronic questionnaire exceeds the quality paper questionnaire because it is controlled by a set of specific rules and regulations that mainly ensure that the data is complete and consistent.
- All data entered stay saved on the devices, in addition to that, a backup copy of data is saved on an external memory added to the device.
- Data is electronically monitored on a daily basis by following up on the enumerators and identifying their level of achievement.
- The confidentiality of data is preserved as logging into the software is possible through entering a password that is exclusively known to the persons authorized for using the software.
- An electronic review is performed, thus reducing disk work by applying many validation rules, and electronic coding systems.

## 4. Data Auditing

Auditing is usually used to detect non-sampling errors, which can be summed up as errors of non-response, interview errors, coding errors, and errors of data entry, in all these sources of errors its measurement along with its impact assessment in terms of the data accuracy is hard.

### 4.1. Types of Auditing

1. Auditing the structure of the questionnaire: This is the stage of auditing the logic sequence of the data included in the questionnaire and its sections; hence, transitions within the questionnaire questions should be considered while excluding the presence of data in case of a certain transition, and vice versa.
2. Validating Data: Validating data through the following:
   - Ensuring that questions with numerical answers do not include any non-numerical characters;
   - Ensuring that there are no missing data for each question requiring data.
   - Ensuring that the answers for each question fall within the right scope.
3. Data Consistency: The term "data consistency" indicates the existence of interrelations between answers within the same questionnaire; such relation can be a logical relation between questions of the same section and the various sections, or it can be a structural relation between the questions, In addition to consistency with the historical data .
4. Cross Tabulations: Regarding the distribution of data, some interrelated tables are established with logical consistency and interrelation of data, in which some extreme values are identified, such as establishing a cross tabulation between age and academic stage while observing the presence of any extreme values of age in some academic stages.

### 4.2. General Guidelines on Data Auditing

The following provides some instructions for auditing data:

- Auditing rules shall be set by the specialized team who has a wide experience within this field as well as in terms of designing the questionnaires, analyzing data, and carrying out similar surveys.
- Auditing shall be carried out at each stage, whereby it does not contradict with the auditing process performed at another stage (The auditing processes carried out while collecting and processing data shall be aligned together).
- Auditing shall aim at providing information, whereby such information is provided in the form of the quality criteria pertaining to the current survey or for suggesting improvements for the prospective survey processes.
- Quality control and assurance processes shall be carried out to limit and address errors that must be notified during the auditing process.

## 5. Non-Response Treatment

Non-response cases differ in terms of an item or a variable, whereby they might relate to a single non-response case pertaining to one variable or item, or they might pertain to many variables, thus being a case of multivariate. On the other hand, non-response cases might be classified pursuant to the processing method applied for missing values, whereby there are deterministic means applied for compensating for all missing values which share similar properties, by the mean value of the data or by median,, and there are stochastic methods applied for compensating for missing values which have not necessarily be similar.

***- Simple Methods for Non-Response Treatment***

In terms of the statistical theory, several traditional methods may be applied for compensating for missing values, such as the Deductive Method which applies logic relationships linking between the value of a variable featuring a missing value and another variable, like the case of linking an academic qualification to the age and years of study. Another simple method is applied for treating missing values, whereby such method replaces all the missing values, such as the average value or median value. This method may further be developed through dividing or stratifying the population into post-stratification layers pursuant to an assisting variable that may be strongly related to the required variable for compensating for the missing values. Then, the missing value is determined, and the average value is bound to the limits of its relevant part, whereby imputation is carried out in this respect. For instance, households can be divided as per the size of the household (number of individuals) for applying such division to compensate for the lost values in terms of the household expenditure variable. Hence, if the household did not respond or provide the expenditure data, and if it included more than eight members, the average or medium rate is calculated for all the households of the same size, and the result shall compensate for the missing values.

Although these methods are easy and simple in treating missing values and used in terms of the various social and economic surveys, whereby they provide relatively unbiased and accurate results, other methods are more advanced and complicated.

***- Linear Regression Method***

This method depends on the availability of auxiliary variables in which data is complete. A linear regression model is elaborated based upon these variables, whereby the variable featuring missing data is a dependent variable, whereas the auxiliary data are independent variables. The accuracy of the imputation results obtained from applying this method mainly depends on the closeness and the correlation between independent variables and the dependent variable. However, in many instances, using this method for processing the missing data distorts the statistical distribution of the remaining data, for example in case the academic years variable of the head of the household and the number of offspring within a household was adopted along with other variables for the elaboration of linear regression model for estimating the age of the head of the household and applying such values for processing the missing ages of some heads of household, such instance might cause a bias in terms of the statistical distribution of the age variation which is supposed to be a normal distribution.

***- Cold Deck Imputation Method***

This method depends on the formula for imputation for the missing values in terms of the current survey adopting the values obtained from a similar survey that was previously conducted for the same population. Upon assuming that a household is formed of four members, whereby only one person was encompassed by the data of the current survey, in case a previously conducted survey for the same population has provided income indicators as per the variables about the number of individuals and the number of workers within the same household, we may substitute the income value for the same variables in the previous survey. In this respect, we may also select one response randomly or adopt the average of responses. However, adopting this method features an inconvenience, whereby the survey's circumstances might defer from one period to another. Hence, accepting this solution would infer that the circumstances are constant.

***- Hot Deck Imputation Method***

This method is based upon the aforementioned method. However, the imputation is carried out pursuant to the data obtained from the current survey rather than from the previous one. Hence, this method is more realistic and may be used in various censuses and surveys.

This method is carried out pursuant to the following steps:

- The application of the stratified classification upon dividing the set of data into categories;
- The selection of a registered value in a category for replacing the missing value in the same category.

Pursuant to the aforementioned, we would notice that there are methods to be applied for compensating or replacing lost values in terms of statistical surveys and that some of these methods are simple. At the same time, some others require in-depth statistical analysis that would link various variables in terms of linear regression relationships. Moreover, selecting the most appropriate method would be relevant to the type of the variable, whereby there are continuous variables, others would require answering them through nominal, ordinal, categorical, or dual variables, such as gender, social status, and relationship to the head of the household.

# 7 | Statistical Estimation

## 1. Introduction

Statistical estimation is the process of elaborating statistical indicators from the available data. It aims to develop statistical estimates (indicators) to use them to infer unknown properties of the population to which the statistical information is relevant, namely in cases where this property is of an unknown value. Such statistical estimates are of various types, such as average, mean, rate, and total.

Developing statistical estimates from survey data can be done in two ways. In terms of the first method, weights are adopted in case data pertains to a statistical sample representing a population, whereby each of the sampling units represents a weight of many population units. As for the second method, it is applied whenever the sample is not weighted. Thus, the relative importance of all sampling units in the sample are same.. Hence, the aforementioned would lead us to draw upon the sampling weights concept to develop statistical estimates or indicators.

**Sampling Weights**

The weighting process relates to calculating the sampling weights ( raising factors) to construct the estimates about a population upon using the results drawn from the survey.. Accordingly, this chapter will tackle the following:

- Types of weights;
- weights adjustment as per the non-response ratio;
- Using additional information for weight adjustments;
- Post-stratification method for weight adjustments;
- Relative estimation; and
- Simple, average, total, and ratio estimates.

## 2. Weighting Sample Results

The weighting process occurs during the estimation process and is the essential part thereof, where each responding sampling unit is assigned a weight. The weight can be considered as a numerical factor calculated as the average of dividing the number of units in the population by the corresponding number of sampling units so that each sampling unit has a certain weight. The weight (Wd) of a sampling unit is the inverse of the selection probability of this unit (π). For example, a sample designed in two stages of selection would include two sampling probabilities: the first is calculated from selecting the unit in the first sampling stage (π1), and the second sampling stage's unit would be (π2). Hence the sample design weight would be calculated as follows:

$$W_d = \frac{1}{\pi_1} \times \frac{1}{\pi_2}$$

For example, if the probability for selecting the sampling unit from the population is 1/ 50, the weight of this unit would then be 50. In other words, this sampling unit represents 50 units in the population.

### 2.1. Types of Weights

Developing weights depends upon the sampling design. Some samples are self-weighting while others are not. Self-weighting samples are those samples that grant equal probability for all sampling units.

- **Weighting for Equal Probability Sample Designs:** When the weights are equal probability for all sampling units, this design is then called a self-weighing sample, i.e., each unit has the same probability of appearing in the sample. Weights can be ignored when calculating some statistics, such as ratios and averages. However, when there is a need for adjustment cases of non-response, weights must be used.

  A self-weighting sample can be used in various designs, such as simple random sampling, systematic sampling, and in some cases of stratified sampling.

  When the sample is drawn using random sampling technique and the sample size of each stratum is proportional to the population size in this stratum, the ratio of the sample to the population (sampling fraction) would be equal for each stratum, and each unit in the population would have an equal probability of appearing in the sample irrespective of the stratum to which it belongs; in this case, all sampling units would have the same weights or importance.

**Example:** Distribution of a sample proportional to size using the stratified random sample technique

Assume that a population of N = 1000 individuals is divided into two strata within the survey frame; the first stratum includes N1= 400 males, the second stratum includes N2 = 600 females. The sample to be selected should have a total size of n=250; this sample is to be proportionally distributed amongst each of these strata according to the stratum size. Hence, the sampling fraction for each class can be illustrated as follows:

**Table (1):** Distribution of the sample proportionally to size using stratified random sample technique

| Sampling Fraction | Sample Size | Population Size | Strata |
|---|---|---|---|
| $\pi 1 = \frac{n1}{N1} = \frac{100}{400} = \frac{1}{4}$ | *n1*=100 | *N1*=400 | Male |
| $\pi 2 = \frac{n2}{N2} = \frac{150}{600} = \frac{1}{4}$ | *n2*=150 | *N2*=600 | Female |
| $\pi = \frac{n}{N} = \frac{250}{1000} = \frac{1}{4}$ | ***n*=250** | ***N*=1000** | **Total** |

As illustrated in the table (1), the sample is distributed proportionally with each stratum size; it is the value of the population size in stratum divided by the population size summed over strata, N, multiplying by the total sample size.

$$ni = \frac{Ni}{N} * n$$

**Whereby:**

- ni refers to the sample size in stratum i
- Ni refers to the population size in stratum i
- N refers to the total population size

The sample size in the first stratum is $\frac{400}{1000} * 250 = 100$ and the sample size in the second stratum is $\frac{600}{1000} * 250 = 150$, whereas the value of the sample fraction for each stratum is calculated separately by dividing the sample size in stratum I by the population size in stratum i, as illustrated below:

$$\pi i = \frac{ni}{Ni}$$

**Whereby:**

- ni refers to the sample size in stratum i
- Ni refers to the population size in stratum i
- πi refers to the selection probability for stratum i.

**Weighting for Unequal Probability Sample Designs**

Despite the simplicity of the self-weighting sample design, it cannot always be implemented, such as in the cases of using a stratified sample design for carrying out surveys in the Emirate, proportional distribution of the samples across strata might lead to sample size in some strata that are too small for analysis purposes. Therefore, there is sometimes a need to select a larger sample size in some areas, although such samples are not needed for large areas.

The following example illustrates designing weights for a stratified sample, where the sample size in each stratum is not proportional to the population size within the stratum.

Example: Distribution of a stratified random sample in a way that is not proportional to the size

In conducting a survey on a private school including 1500 students distributed among two sections, the size of the population in section A is N1= 900, whereas the population size in section B is N2= 600. A sample of n1=200 students has been selected in stratum A, and a sample of n2=250 students has been taken from stratum B. What are the weights of the sample chosen?

**Table (2):** Distribution of a non-proportional sample to the size using the technique of stratified random sample

| Sampling Fraction | Sample Size | Population Size | Strata |
|---|---|---|---|
| $\pi 1 = \frac{n1}{N1} = \frac{50}{500} = \frac{1}{10}$ | *n1*=50 | *N1*=500 | Section A |
| $\pi 2 = \frac{n1}{N1} = \frac{200}{1000} = \frac{1}{5}$ | *n2*=200 | *N2*=1000 | Section B |
| $\boldsymbol{\pi = \frac{n}{N} = \frac{250}{1500} = \frac{1}{6}}$ | ***n*=250** | ***N* =1,500** | **Total** |

As illustrated in the table (2), the sample is distributed un proportionally with each stratum size.

$$ni = \frac{N\text{i}}{N} * n$$

**Whereby:**

- ni1 refers to the sample size in stratum i
- Ni1 refers to the population size in stratum i
- n refers to the total sample size in all the strata
- N refers to the total population size in all the strata

The value of the sampling fraction in the first stratum, "section A" is 1/ 10, whereas the value of sampling fraction in the second stratum "section B" is 1/ 5. Therefore, the value of each respondent's weight under the section "A" stratum is Wd = 10, whereas the value of each respondent's weight under the section "B" stratum is Wd = 5.

### 2.2. Weights Adjustments as per the Non-Response Ratio

Different response rates are achieved across various surveys. Non-response occurs when some or all of the required information from the sampling units is not obtained for whatever reason. Cases of non-response are divided into two types:

First Type: Non-response of the entire sampling unit, such as the non-response of a whole household to a specific survey.

Second Type: Partial non-response, such as the cases where the sampling unit can partially respond to some questions of the questionnaire, not to others.

In cases of non-response of the second type, partial non-response, the most popular processing would be the imputation technique for the missing values. However, in the first type, when the non-response includes the entire sampling unit, it is preferable to resort to a weight adjustment technique. The adjusted weight can be expressed through the following equation:

$$W_{adj} = W \times W_r$$

**Whereby:**

- $W_{adj}$ : refers to the adjusted weight;
- W : refers to the initial weight
- $W_r$ : refers to the inverse of response rate.

In other words, the non-response cases can be used to adjust the sampling weights where the sampling weight of non-responding units is proportionally distributed amongst the rest of the sampling units that have responded.

### 2.3. Using Additional Information for Weight adjustment

Sometimes, the weighting process does not end at calculating and adjusting the weights according to non-response. There might be other possibilities for improving the weights. For example, when data for the population is available from other sources (such as a census), this information can be used to adjust the weights to reflect the population size by redistributing the non-response adjusted design weights in the sample proportional to the population values.

Reasons for using additional information for weight adjustments:

- **First,** it is important to match the survey estimates with the groups of the population or estimates from other sources for the survey results to be more accurate. For example, in many social surveys, the survey estimates are adjusted to be consistent with the census estimates (age, gender distribution, etc.), and additional information can be used from data of administrative records or previous surveys of high accuracy.
- **Second,** using additional information contributes to improving the accuracy of the survey estimates, whereby an estimate is considered of high accuracy when the sampling error is small. Hence, it is preferable to use additional information when designing the sample so that it contributes to increasing the efficiency of the resulting estimates.

### 2.4. Post-Stratification Method for Weight adjustments

The post-stratification method is used to adjust the survey weights upon using a variable appropriate to the stratification technique. However, this method cannot be applied during the design stage because data is not available or updated, whereby stratified information of a population is usually more accurate and reliable after the sample is selected.

The post-stratification can be used when the additional data are of continuous type, for example, the number of men and women in the population. This method is considered one of the most effective methods for diminishing the sample variation when the differences between the sample weighted estimates and their corresponding population distributions for the post-stratification data are relatively large. Nonetheless, it is worth noting that stratifying at the design stage is more accurate than post-stratification.

# 3. Statistical Estimation

### 3.1. Ratio Estimates

Ratio estimation is one of the most popular methods when using additional information to improve the survey estimates, namely when data is continuous , whereby Ratio estimates would be similar to the post-stratification method. In ratio estimation, weights are adjusted in the different strata using a multiplier factor which is the estimation drawn from the additional variable to the corresponding value for the sample for cases in the same class group.

For example, if the survey objective is to estimate the total agricultural area for the wheat crop in a certain region, the information on the total agricultural area in each region is used as a piece of additional information. If the total agricultural area of the wheat crop is greatly correlated with the total agricultural area, this information could be used to improve the estimates of the total agricultural area of the wheat crop. Hence, upon applying the Ratio estimation method, the modification factor applied for each region would be reached by multiplying the ratio of wheat agricultural area in the sample by the total agriculture area in the region as a whole.

### 3.2. Simple Estimates

***- Simple Estimates for Unweighted Data***

Unweighted estimates for the size or number of units in the sample should not be used as an estimate for the size of the population, as it is merely an expression of the total sample. Neither is it possible to rely on the total of the sample units to estimate the total population using unweighted data, as there is a clear difference between the total population and the total sample. As for the arithmetic mean of the unweighted data, it can be used to estimate the population's mean, though it may be of less accuracy and efficiency than it would be if the sample data were weighted.

- Estimating the total of the sample:

$$\hat{y} = \sum_{i \in S_r} y_i$$

- Estimating the mean of the sample:

$$\hat{\bar{y}} = \frac{\sum_{i \in S_r} y_i}{n}$$

Whereby n represents the sample size, being the simple estimates for data.

As aforementioned, simple descriptive statistics, just like totals, means and ratios are calculated for each survey, in the case of quantitative data which are relevant to interval or continuous variables.

***- Simple Estimates of Weighted Data***

Estimating the Population Total

$$\hat{Y} = \sum_{i \in S_r} w_i \, y_i$$

**Whereby:**

- i refers to the responding unit of the sample;
- wi refers to the final adjusted weight of unit i;
- Yi refers to the value of unit i; and
- Sr refers to the group of responding units.

Example: In expenditure data, the total value can be calculated, such as the total expenditure of the households in the population, whereby it is the product of multiplying the final weight of each sample household by the value of expenditure of each responding household in the survey.

***Estimating the Mean of the Population***

The value of arithmetic mean for quantitative data in a population is the sum of the product of the value by the related weight of each sampling unit, whereby the summed product is then divided by the total weight for all sampling units.

$$\hat{\bar{Y}} = \frac{\sum_{i \in S_r} w_i \, y_i}{\sum_{i \in S_r} w_i} = \frac{\hat{Y}}{\hat{N}}$$

Whereby $\hat{N}$ refers to the total weights.

***relative Estimation***

The relative of units in the survey population to a specific variable is estimated as the total weights of the units to which this variable applies, divided by the total weights for all sampling units.

$$\hat{P} = \frac{\sum_{i \in S_r \cap C} w_i}{\sum_{i \in S_r} w_i} = \frac{\hat{N}_c}{\hat{N}}$$

Whereby C refers to the total units to which the variable applies.

Example: In a household survey, the percentage of males in the population is the total sampling weights of males included in the sample divided by the total sampling weights for all sampling units, being both males and females.

Often, when calculating statistical estimates for the results of a survey, the estimation process is not only restricted to the population level as a whole but is often required to be carried out on sub or partial levels, such as in the cases pertaining to estimating the average expenditure of an individual as per their age group category, or estimating the average size of a household according to a specific geographic region. The size of the population is estimated, as illustrated below:

$$\hat{N}_{domain} = \sum_{i = S_r \cap domain} w_i$$

Estimating the Total for a variable Y is given by:

$$\hat{Y}_{domain} = \sum_{i = S_r \cap domain} w_i \, y_i$$

Estimating the Average of the variable Y. is given by:

$$\hat{\bar{Y}}_{domain} = \frac{\sum_{i = S_r \cap domain} w_i \, y_i}{\sum_{i = S_r \cap domain} w_i} = \frac{\hat{Y}_{domain}}{\hat{N}_{domain}}$$

# 8 | Dissemination of Statistical Data

## 1. Introduction

The statistical report of the data results is a structural form for the presentation of data and information. As such data assume a key role in terms of the decision-making process, the techniques for elaborating reports are essential. Reports are also considered as a means of communication between the various units within the same institution or between the different institutions and departments. Thus, the preparation of the reports is considered a technical process that requires high technical skills and experience and demands tremendous efforts to ensure high quality, mainly due to the multiple aspects it involves.

The definition of a report here refers to a written presentation of data, information, and facts related to a specific topic or issue, in the form of an analytical presentation supported by statistical facts which are presented in a simple, organized way, whereby the report also includes the relevant findings and suggestions. Moreover, since reports often represent official documents that others read, such reports shall always be concise, clear, and accurate.

Some researchers find it hard to separate between data and information, as data refers to facts, numbers, letters, words expressing an idea, an issue, or a specific situation, such as sales figures, production figures, or the number of workers, among others; while information refers to organized facts that directly contribute to the decision-making process. This information also relates to the findings of the data describing the events about operations, such as the inflation or unemployment rate, among others, whereby they are obtained in a meaningful form for the beneficiaries.

The statistical data report is of great importance because it represents the final statistical product that will be presented to the users through its dissemination either through the press, websites, or interviews, among others.

## 2. Survey Findings Main Report

The main survey report is considered one of the essential products of the survey. It is usually the first statistical report to be made available to users. Hence it includes much information on the objectives and methodology of the survey and the documentation of the concepts, definitions, key findings, and tables.

Consistency in the structure of the survey report helps the users obtain statistical data and information on the survey. Most probably, government entities and statistical agencies have defined policies and guiding standards and principles regarding the content and consistency to ensure the production of this report per the users' requirements. Moreover, clear controls should be set to develop the report to follow the same rules in terms of formatting. The following provides a suggested structure for the survey report, which is applicable in most cases:

- Cover page: Usually, it is a standardized design, including the logo of the government entity and identifying the topic and reference period.
- Table of contents: Helps the user to find the information easily.
- List of tables and graphs: Many users prefer a visual presentation of survey findings.
- Executive summary: It might be optional, but it is highly recommended, as it helps in highlighting the conclusions of the most important survey findings.
- Introduction: A short and general brief that provides the main information related to developing the survey, such as the date of the project, reference period, sponsors, the study objectives, a quick brief on the methodology and the purpose of the report, in addition to providing a general view on the sections included in the report and their correlations.

- Paragraphs: The report is divided into several paragraphs that include all the details on the main conclusions and the presentation of findings followed by more detailed information so that ideas are arranged in a logical sequence based on their importance. It includes some tables and graphs alongside an explanation of the meanings of these results and their importance in supporting the report.
- Conclusion: This section presents the conclusions on the necessary procedures to be applied by the key decision-makers and users, whereby such conclusions and recommendations are adopted by the analysis specialists and experts.
- Recommendations: This section is optional in technical reports. Recommendations could be issued regarding solutions for the problems that appeared during any stage of the survey. This is usually done in order to improve surveys that might be facing the same situation.
- List of references: It includes all references used as a guide in developing this report.
- Contacts: This section is mandatory in each survey as it provides the direct contact number to facilitate communication for users and refers to the website of the statistical entity where information could be found.
- Appendix: The purpose is to provide a brief and clear idea of the main details supporting the text.

## 3. Guide for Writing the Survey Findings Main Report

Upon publishing statistical information, the content should be consistent with the standards and controls included in the guide. If the content of the report is not compatible with these standards, it cannot be published and would, thus, be subject to correction to ensure it is abiding by all the required standards and controls. To ensure that all statistical publications are consistent and efficient and that repetition is avoided as much as possible, the following provides some guidelines for writing statistical reports:

**Clear Language:** Sentences should be clearly formed, appropriate words and linguistic expressions should be used according to the topics, unnecessary detail should be avoided to avoid confusion and ambiguity. It is important to avoid exaggerating in both summarizing and elaborating on sentences in order to clearly deliver the meaning to the reader.

**Abbreviations and Initials:** These should be used moderately; terms are to be defined in full when first mentioned in the report so that the reader will understand its meanings, to enable the use of abbreviations in further parts of the text, such as Consumer Price Index (CPI).

**Consistent Terms:** Terms should be fixed and consistent throughout the whole report. For instance, when the title and the text refer to assets and liabilities, it would not be acceptable to use the title as "Balance Sheet Table or Graph." Categories should be either identified or clearly mentioned, ensuring that the word "other" is not used to avoid ambiguity.

**Dates:** Using ambiguous expressions to refer to dates, as in "last year" or "last month," is not acceptable, whereby the month and year should be identified, such as "July 2005".

**Reference Period:** Reference period should be immediately inserted below the main title upon publishing the reports or the primary data, such as noting the following: "Monthly Survey of the Household Income – December 2002". Moreover, if the report is related to preliminary data, the latter shall be mentioned to avoid redundancy within a text.

**Main Title and Subtitles:** All publications titles are to be clear and concise, while merely mentioning the topic titles and the reference period so that the main title would refer to the dates and base years, such as mentioning: "Real Gross Domestic Production at the Cost of Production Factors by Industry – July 1993 (Primary Data)". Moreover, the used subtitles shall be short and simple to guide the reader to the information included in the text.

**Content Editing Standards**

Publications should include the following elements:

- Title page;
- Preface, if required, in case the publication is new or of extreme importance;
- Table of contents, which is automatically elaborated by Microsoft Word;
- Introduction;
- Main points, which should include the summarized graphs and tables;
- Statistical tables;
- Annexes, which should include an index of the abbreviations, the sources of data, used methodologies, technical notes, the prospective date of issue, and any other additional information.

## 4. Statistical Tables

Tables are used when information, generally being statistical information, is to be delivered in a better way, whereby such way would be providing information in the form of tables rather than in the form of texts. Tables are used to illustrate or study quantitative data, making it easier to detect relations between several variables, thus allowing for direct comparison.

Standard technical procedures are adopted for elaborating statistical tables to guarantee consistency amongst all statistical reports and bulletins. In practice, the following are general guidelines on the elaboration of statistical tables, whereby such guidelines can be followed by other entities, and relate to the following:

- The fond to be used is to be identified for tables, whether in English or Arabic.
- It is recommended to use a 10-point size for the table contents, no less than a font of eight points, and no more than a 12-point space for the title.
- A line spacing of 0.5 shall be used for the horizontal line under the title and the column titles, whereby there is no need to use vertical lines.
- The same font size shall be used for the column titles and words, main points, and table contents. Usually, the size used in this respect shall be smaller by two sizes than the line spacing used for the publication text (no less than 8 points).
- Footnotes (margin notes) shall have a line spacing of 1 to 2 points less than that of the main table content.
- Tables, if any, shall be numbered, whereby numbering can be as follows: 1,2,3,4, etc., or 1.1, 1.2, 1.3., etc.
- The "Table Title" option shall be used to allow the automatic arrangement of the list of tables.
- In general, the same width is to be used for all data columns.
- Data sources are to be mentioned all the time, such as mentioning the following: Statistics Centre – Abu Dhabi.

## 5. Graphs

Most of the time, graphs reflect information better and quicker than texts and tables. Yet, there are other ways to communicate information, such as schemes and maps. Moreover, there are various types of graphs of which the most used are charts and column graphs. It is preferable to use these two simple types of graphs. Generally, three-dimensional and complex graphs are not to be used because they are too difficult to understand. General guidelines:

- Detailed data of graphs are to be made available in the attached tables (either accompanying the graph or the graph's location is referred to in the publication).
- The font and the font size shall be set for both Arabic and English texts.
- There is no need to place the graph in a box.
- To mention various periods in the table, the whole period name is not to be mentioned, whereby mentioning the month shall be sufficient.
- The font size used is to change according to the nature of the text and the type of arrangements used within.
- It is recommended that graphs and schemes used within the statistical publication are numbered as 1, 2, 3, etc., or 1.1, 1.2, 1., etc.
- Table key is to be used in case the graph represents two or more groups of data such as data related to the Emirate and data related to regions abroad. Usually, the key for the data included in the graph is placed between the graph title and its content.
- The type of graph to be used shall be identified.
- Chart graphs are considered an effective tool to reflect changing trends in the data across different times in an expressive, visual way. Thus, it is the ideal type of graph to visualize time series.
- It is possible to use Bar chart to compare different groups of data or compare the same group of data in two or more time periods.
- pyramid graphs, are not to be used except in terms of reflecting the age/gender composition.
- Pie charts are used as an alternative for chart graphs or column graphs to highlight, for instance, the size of governmental expenditure layers based upon the type and purpose. Pie charts have to be used with great care, though, as it may lead to wrong interpretation of data, especially if there were some sectors with relatively equal size.

## 6. Data Quality Report

Assessing data quality includes the assessment of the final deliverables of the statistical process, namely in terms of achieving the objectives, the efficiency and accuracy of the data produced by the statistical operation, and the contribution of such data to the development of the future survey's data. On the other hand, this process allows data users to develop interpretations and analyze the survey data.

To boost the confidence of the users of statistical publications related to the entity, developing a summary of the quality of the published data requires an explanation and illustration of all aspects related to quality, whereby it would allow data users to use the statistics appropriately, as users need information on the survey design, data collection, and the adopted methodology and processes, as well as other factors that impact the error size, whereby it is possible to identify a number of the key standards which indicate the quality of the final results of the survey.

It is also possible to include a comprehensive quality report in the statistical publications, whereby such a report would be comprehensive and assess data quality and serve as quantitative criteria for measuring the quality of statistical data. Hence, it is essential to include various statistical criteria and standards about the quality of data and its emanating indicators, whenever possible, whereby some of such standards include the following, among others:

- Standard errors and confidence intervals
- Coefficient of variation;
- Design Effect;
- Eligible Rate of Sampling Units;
- Response and Refusal Rates;
- Editing Rate;
- Average Time Need for Filling the questionnaire.

## 7. Confidentiality Procedures and Survey Finalization

This section relates to developing the disclaimer policy and terms of using the website and highlighting the rights and responsibilities of the party responsible for data dissemination. These policies are used to protect the entity against any external claims and protect the statistical entity against any legal liabilities that might arise from copying or misusing information produced by the entity. Unless otherwise indicated, statistical survey publications or those produced by the entity should be protected by copyrights and publication rights. Copying of the report contents (tables and graphs, etc.), in part or in full, by all means, shall be carried out according to obtaining special authorization from the entity, given it is acknowledged that the approval is issued by the entity, upon mentioning the name of the entity, the publication year, the indexing number, the reference period, and the page number.

## References


Survey Methods and Practices, Statistics Canada, 2010.

Statistical Sampling Guide, Statistics Centre – Abu Dhabi, 2015.

Statistical Terms Dictionary, Sampling Terms 2005, Arab Institute for Training and Statistical Research.

Editing Statistical Publications Guide, Code RE-DA-ST-DI-001, Version 1, Revision 1.

Disclaimer and Terms of Use, Code PO-DA-DIS-001, Version 1, Revision 0.

Statistical Data Quality Framework, Statistics Centre – Abu Dhabi, Code RE-DA-SPD- 001.

Researchers Guide in Statistical Analysis, Aida Rizkallah, 2006.

Research Methods for Managers, Uma Sekaran, 2006.

Scientific Research Methodologies, Mohammad Obeidat et Al., Amman, Wael Publishing House, 1999.

Guiding Principles for Developing and Implementing Statistical Surveys, Guide no. (2), National Statistics Center, December 2011.

Statistical Terms Dictionary, Sampling Terms 2005, Arab Institute for Training and Statistical Research.

Sampling Methods Handbook- Cochrane, 1977.

Survey Methods and Practices, Statistics Canada, October 2003.

Fritz Scheuren: What is Survey? July 13, 2004.

Essentials of Research Design and Methodology, Geoffrey Marczyk & others, 2005.

The Power of Survey Design, Giuseppe Larossi, 2006.

مـركـز الإحـصـاء
STATISTICS CENTRE

**الرؤية:** ببياناتنا نمضي نحو غدٍ أفضل
**Vision:** Driven by data for a better tomorrow

www.scad.gov.ae
adstatistics