



Opinion Polls Implementation Guide

Methodology and Quality Guides - Guide No. (11)



Table of Contents

Introduction	3
Chapter One Introduction to Opinion Polls	4
Chapter Two Stages for Implementing Public Opinion Polls	5
Chapter Three Opinion Polls Sampling Methods	25
Chapter Four Opinion Polls Data Quality	29
References	36

Introduction

Within the scope of its work, the Statistics Centre – Abu Dhabi elaborates guides and sets the guidelines related to the statistical processes it carries out, including statistical surveys and opinion polls, among others, as to inform data users of all categories of the various work methodologies related to data collection, processing, control, as well as data retrieval and analysis.

In accordance with the aforementioned, this guide includes four chapters, the first of which serves as an introduction to opinion polls and provides a general overview of the history, importance, objectives, and fields of implementation of opinion polls, whereas the second chapter tackles the methods and stages of implementing opinion polls, as of the preparatory stage up to drawing findings. As for the third chapter, it relates to the statistical sampling methods in terms of the design and drawing of sampling units, and covers the key principles of statistical analysis, including the application of descriptive statistics through graphs and tables, the calculation of some indicators, the application of statistical inference, involving statistical testing hypothesis related to one statistical population or two separate ones, and tackles the Chi-squared test for identifying the autonomy of population. Finally, the fourth chapter presents the principles and procedures for verifying the quality of statistical operations related to carrying out opinion polls.

It shall be noted, in this respect, that other detailed statistical guides are issued by the Statistics Center – Abu Dhabi, whereby such manuals, namely the guides for statistical analysis principles, statistical sampling, statistical surveys implementation, and statistical data quality, are considered as a reference frame for this guide.

1 | Introduction to Opinion Polls

Opinion polls are a technical method for collecting information used for obtaining the opinion of a group of individuals, on a given topic, at a specific time and place.

A public opinion poll and taking its indicators into account form a key basis for guiding decision-makers in terms of all the economic and social fields toward the right direction.

Opinion polls include the measurement of the public's expectations, the assessment of the governments' performance, carrying out scientific studies on local issues which fall within the scope of interest of the Emirate of Abu Dhabi, whereby such studies are carried out in an accurate, objective, and neutral manner, and providing such information to researchers and decision-makers as to foster their ability to make decisions scientifically and objectively.

Accordingly, opinion polls have established themselves as an important factor for measuring public opinion in the Emirate of Abu Dhabi and as one of the most important bases for decision-makers in various fields. Hence, there is no doubt that achieving success is subject to the measurement of the opinion of the community's individuals and institutions, as being aware of their opinion contributes to ensuring information flow to decision-makers and community members. Furthermore, opinion polls also raise awareness among the concerned persons, thus achieving a positive indicator contribution to making the right decision.

Fields of use of opinion polls

Opinion polls are implemented upon adopting the relevant scientific approach and applying the best relevant practices and techniques to obtain accurate and reliable information, which may be planned upon in terms of steering the necessary policies aimed at identifying any change in terms of public opinion related to the tacked fields.

The objectives of opinion polls are many and involve many fields, whereby public opinion polls relate to opinions about, economic, or social issues, whereas other polls are implemented to achieve scientific purposes, such as investigative studies, which provide important information for the various research fields and consist of many types, such as marketing research, psychology, sociology, and health fields, among others.

Furthermore, there are many situations where opinion polls are implemented, such as the following:

- Social and Economic Surveys: These polls tackle the economic and social status of the various community groups as well as the forecasts pertaining to their economic situation in comparison with previous or current periods, like the Consumer Confidence Index Poll and the Business Climate index.
- Satisfaction / Happiness Assessment Polls, including:
 - Customer Satisfaction / Happiness Measurement;
 - Employees Satisfaction / Happiness Measurement;
 - Suppliers Satisfaction / Happiness Measurement;
 - Partners Satisfaction / Happiness Measurement; and
 - Community Satisfaction / Happiness Measurement in terms of the services they are provided with, such as the community satisfaction in terms of the roads' cleanliness and lighting, public parks, sports facilities, pedestrian crossings, and water quality, among others.

2 | Stages for Implementing Public Opinion Polls

In order to implement opinion polls carried out based upon a correct scientific methodology, this guide has adopted the best international practices pertaining to opinion polls and the methodologies adopted in terms of statistical operations, pursuant to the Generic Statistical Business Process Model (GSBPM), whereby this guide shall apply to all the opinion polls implemented by the Statistics Centre – Abu Dhabi.

Furthermore, this guide also depends on summarizing the stage of implementing opinion polls and limiting them to four stages, the first of which is the planning stage, followed by the preparation and implementation stage, then the findings drawing stage, as shown below:

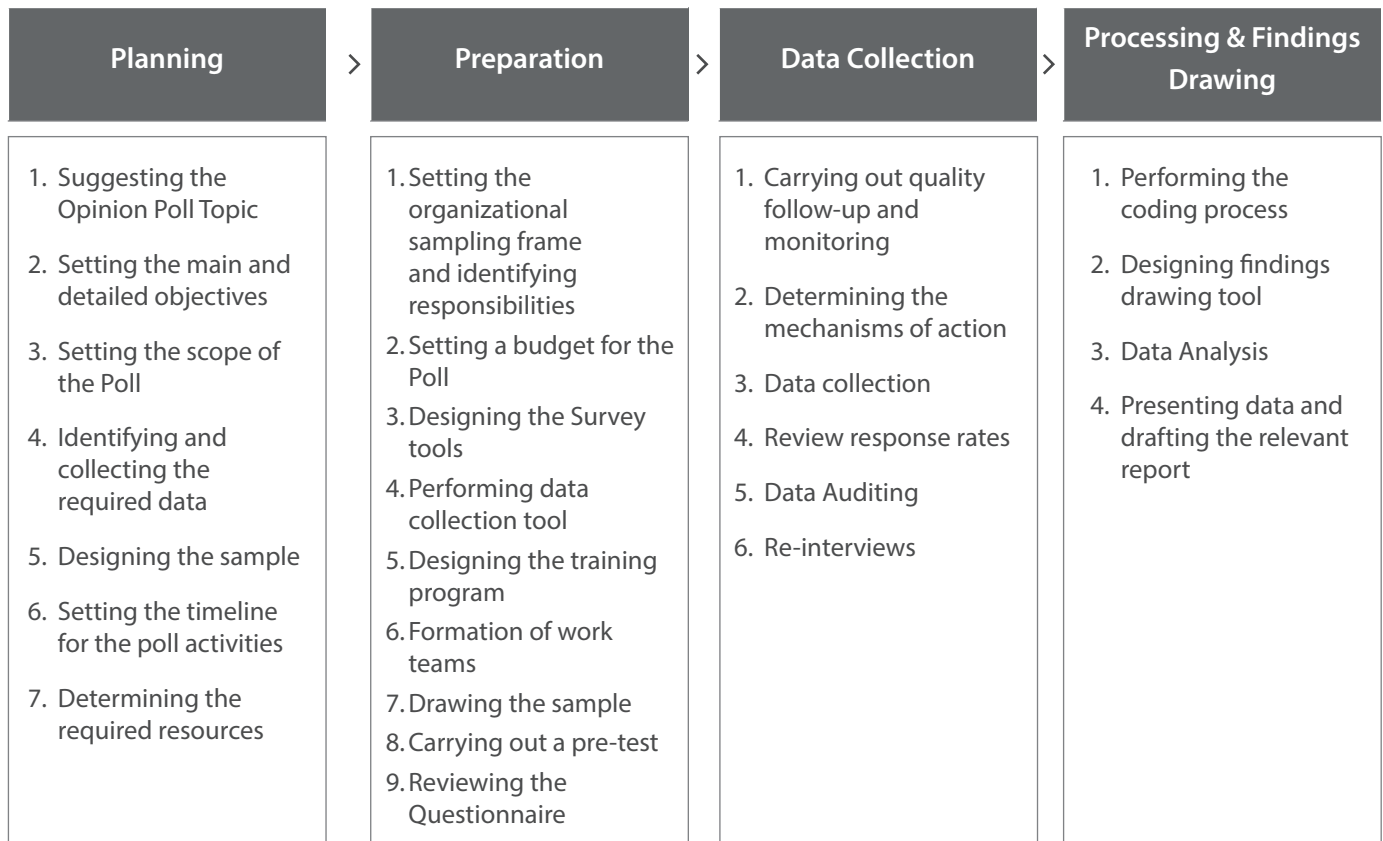


Figure (1) – Stages for Implementing Opinion Polls

2.1. Opinion Polls Planning Stage

There is no doubt that the lack of planning in terms of any work would hinder its success. Hence, if we are to carry out an opinion poll, the planning stage shall be of utmost importance, namely due to the variety of activities, the large number of participants, the pressure exerted for finalizing the Poll as soon as possible, and the presence of some activities which must be respectively achieved, whereby the deliverables of some of these activities are considered to be the inputs required for carrying out another activity. For instance, it is necessary to finalize the sample design prior to drawing it, whereas it is also important to draw the sample prior to collecting data. Accordingly, the most important stages of the planning process include the following:

1. Suggesting the Opinion Poll Topic

Suggesting the opinion poll topic is the first step in planning a new survey by defining the data to be available from the survey and the outputs to be built. The proposal is prepared after a preliminary study of the available data from administrative records or from previous field surveys, and the proposal is prepared after consulting with experts and specialists from various relevant fields to discuss the justifications for the survey and the feasibility of its implementation.

2. Setting the Main and Detailed Objectives

- The objectives of the Poll to be carried out shall be set clearly and understandably.
- The objectives shall be set pursuant to a general objective that identifies the main objective for carrying out the Poll.
- The detailed objectives supporting the general objective are set in a more specific manner.
- Setting the objectives clearly is important as such objectives form the frame of reference to be complied with by the persons in charge of implementing the Poll's activities.

3. Setting the Scope of the Poll

- The scope of the Poll and its size are set pursuant to the available financial and human resources.
- This step is considered as an initial visualization for a number of alternatives that set various scenarios for achieving the objectives of the survey.
- These assumptions are concretized as to submit them before the entity responsible for the Poll in order to assign a preference to one of the alternatives.
- Upon setting alternatives, an estimative cost and the advantages of each of them shall be assigned thereto, namely in terms of the possibility of generalizing findings and the ability of each alternative to avoid bias in terms of such findings, whereby it shall also be clarified that such advantages would not be considered as scientific affluence. Rather, they would form the cornerstone for elaborating successful policies and designing efficient programs. Hence, upon setting the scope of the Poll, the latter shall take the size of the sample, the level of data dissemination (coverage), and cost into account.

4. Identifying and Collecting the Required Data

- Identifying the Required Data:

This stage relates to identifying the number of variables which would later be reflected in the questions listed under the questionnaire, whereby the questionnaire must meet the needs of the other institutions in terms of the data relevant to the Poll.

- Identifying the Data Collection Unit:

This step relates to setting the sampling unit, which usually is the "individual of a household" or an "establishment" in terms of economic polls. The unit is set in line with the poll objectives and the type of data to be collected.

5. Designing the Sample

In order to design a sample that represents all the groups of a population to be examined, an updated and comprehensive frame shall be provided for the individuals of the target population to ensure that accurate results are drawn from the survey. Carrying out such steps is subject to the following:

- The availability of a list of the names and addresses of all the individuals of a population, whereby the elements of such list are distributed as per the region, cluster, neighborhood, enumeration area, building number, residential unit, and household number;
- The availability of maps highlighting the locations of the buildings, residential units, and streets, pursuant to the appropriate measurement scale;
- The availability of updated lists, including the names and phone numbers of clients, suppliers, strategic partners, and employees, among others.

It shall be noted, in this respect, that the Third Chapter herein provides details on the sample design procedures, their types, , pursuant to previously provided facts.

6. Setting the Timeline for the Poll Activities

During this stage, computer software is used in terms of setting the timeline for implementing all the poll activities, such as Microsoft Project Management.

7. Determining the Required Resources

During this stage, the following needs are identified:

- The Required Software and Applications.
- The Required Devices: Such as iPad devices, photocopy machines, printers, and computers.
- The Identification of Human Resources: This is carried out pursuant to the type of the Poll.
- The Training Location: Training is provided to the entire team involved in the Poll by using the most recent means and technologies.

2.2. Opinion Polls Preparation Stage

Once the poll preparation stage is completed, it is necessary to elaborate the timeline for implementation, to establish an organizational framework for implementing the Poll, whereby such framework shall list all the responsibilities and roles and determine the poll budget and determine the sampling frame and design the survey's tools, as well as to draw the necessary sample, perform the pre-testing process as to test the survey's tools and verify that all the questions are clear and understandable, to review the compliance and coding rules, as well as all the plans related to human resources, whereby such stage would form a practical training stage for the research team formed of data collectors. Then, the work team is formed, and the study population is prepared by clarifying the Poll objective and importance through the various media.

1. Setting the Organizational Framework and Identifying Responsibilities

During this stage, the organizational framework is established whereby it is relied upon in terms of identifying all the activities and their assignment to the team members and identifying the tasks, duties, and responsibilities assigned to each of the team members. It shall be noted.

2. Setting a Budget for the Poll

During this stage, a detailed budget is set for the Poll, whereby this budget encompasses all the items required for implementing the detailed activities, while we shall be keen not to forget any of such items to avoid any ambiguity at work. Hence, the impacting factors which must be taken into account upon setting the budget include the following:

1. The sample size and its scope of coverage;
2. The type of the targeted category (individuals, households, establishments);
3. The size of the questionnaire;
4. The poll methodology; and
5. The timeline.

3. Designing the Survey Tools (Questionnaire)

During this stage, the method to be applied in terms of collecting data is set. In other terms, shall the Poll be conducted through any of the following?

- Face to face interview through using the CAPI system;
- Phone interviews through using the CATI system;
- Audio self-interview through using the ACASI system;

CAPI: Carrying out Computer-Assisted Personal Interviews through computer devices

CATI: Carrying out Computer-Assisted Telephone Interviews through a dedicated computer system

ACASI: Carrying out Audio Computer-Assisted Self Interviews

- Through the Internet: Sending the link through e-mail or SMS;
- In-depth interview;
- Group interview;
- Focus group;
- Distributing the questionnaire to a group of individuals, such as students, or employees, among others; or
- Other means.

Main Questionnaire Design Rules:

- o The Questionnaire shall meet all the poll objectives: The poll objectives are reviewed and compared with the questions listed under the Questionnaire, whereby a comparison is also carried out between the set objectives and the fact that questions meet such objectives or not.
- o A verification process is carried out to ensure that all the questions were formulated clearly and understandably, whereby each question may not be interpreted in multiple ways.
- o A verification process is carried out to ensure that the questionnaire provides some instruction on completing it, namely since the questionnaire includes some questions which require a different way of providing answers.
- o A verification process is carried out to ensure that the questionnaire does not include any suggestive question, that the questions listed therein avoid embarrassing the respondent, and that they do not consist of provocative questions.
- o A verification process is carried out to ensure that the questions are ordered pursuant to a logical sequence, whereby the questions assist the respondent to focus on them and do not cause his distraction. The questions are to be divided into sections as per the topics covered by the study.
- o The questions of the Poll shall not be many, whereby the questionnaire shall not be long, as to avoid causing boredom for the respondent, whereby the answers may be illogical or inconsistent with one another.
- o The Poll shall include a special section listing the demographic and general properties and characteristics of the respondents, such as their age, gender, academic qualification, expenditure, marital status, and employment status, among others.
- o Other technical matters shall be taken into account upon designing the questionnaire, such as the confidentiality and privacy of data, among others.

Standards Adopted for the Public Opinion Polls and trends

1. Dichotomous: Provides two fully contradicting options, such as the following:

Yes / No;	Beneficial / Non-Beneficial
True / False	Suitable / Unsuitable
Fair / Unfair	I think so / I do not think so
Positive / Negative	I love it / I hate it
Expensive / Cheap	I would do it / I would not do it
Influential / Non-Influential	I want to / I do not want to

2. Measuring the Extent of Agreement: In case the Poll aims at measuring the extent of agreement of the respondents on specific paragraphs, opinions, or programs, the standard to be adopted in this respect would be measuring the extent of agreement, whereby the most famous scale in this respect is the "Likert Scale," which may either be based upon three points or five points, as follows:

Three Points Likert Scale for Measuring the Extent of Agreement:

Agree	Neutral	Disagree
(3)	(2)	(1)

Five Points Likert Scale for Measuring the Extent of Agreement:

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
(5)	(4)	(3)	(2)	(1)

3. Scales Measuring Levels: For instance, in case the Poll is aimed at measuring the degree or level of enablement or loyalty among employees, or the level of work pressure, or the level of prices, the "Likert Scale" may not be used for measuring the extent of acceptance, but other scales based upon three or five points may be adopted in this respect, as follows:

Three-Point Scale for Measuring Levels

High	Moderate	Low
(3)	(2)	(1)

Excellent	Good	Weak
(3)	(2)	(1)

Five-Point Scale for Measuring Levels:

Very High	High	Moderate	Low	Very Low
(5)	(4)	(3)	(2)	(1)

Excellent	Very Good	Good	Acceptable	Weak
(5)	(4)	(3)	(2)	(1)

4. Scales Measuring the Extent of Using Specific Methods or Tools: In case a poll is aimed at identifying the extent of adopting specific methods or tools, the most appropriate scale to be adopted in this respect would be one of the following:

Three-Point Scale for Measuring the Extent of Using Specific Methods or Tools:

High Use	Moderate Use	Low Use or No Use
(3)	(2)	(1)

Five-Point Scale for Measuring the Extent of Using Specific Methods or Tools:

Very High Use	High Use	Moderate Use	Low Use	Very Low Use
(5)	(4)	(3)	(2)	(1)

Three-Point Scale for Measuring the Extent of Using Specific Methods or Tools in Terms of Times of Use:

Always	Sometimes	Never
(3)	(2)	(1)

Five-Point Scale for Measuring the Extent of Using Specific Methods or Tools in Terms of Times of Use:

Always	Often	Sometimes	Rarely	Never
(5)	(4)	(3)	(2)	(1)

5. Scales Measuring the Extent of Availability of Specific Elements or Tools: In case a poll is aimed at identifying the extent of availability of specific elements or tools, the most appropriate scale to be adopted in this respect would be one of the following:

Three-Point Scale for Measuring the Extent of Availability of Specific Elements or Tools:

High Availability	Moderate Availability	Low Availability or Unavailability
(3)	(2)	(1)

Five-Point Scale for Measuring the Extent of Availability of Specific Elements or Tools:

Very High Availability	High Availability	Moderate Availability	Low Availability	Very Low Availability
(5)	(4)	(3)	(2)	(1)

6. Scales Measuring Satisfaction / Happiness: In case a poll is aimed at identifying the extent of satisfaction or happiness in terms of services or other matters, the most appropriate scale to be adopted in this respect would be one of the following:

Three-Point Scale for Measuring Happiness / Satisfaction:

Satisfied	Neutral	Unsatisfied
(3)	(2)	(1)

Five-Point Scale for Measuring Happiness / Satisfaction:

Very Satisfied	Satisfied	Neutral	Unsatisfied	Very Unsatisfied
(5)	(4)	(3)	(2)	(1)

Ten-Point Scale for Measuring Happiness / Satisfaction:

Very Unsatisfied										Very Satisfied
0	1	2	3	4	5	6	7	8	9	10

7. Comparison Scales: They are used for comparing between a situation or status among a specific period and another, such as comparing between the current economic status of a household with its economic status over the past year:

Improved Situation	Constant Situation	Worse Situation
(3)	(2)	(1)

4. Performing Data Collection Tool

- Computer software is designed to ensure that the questionnaire data is inputted through various means or methods, whereby data may be uploaded to iPad devices or desktop computers, software which is accessible through links sent through the website of the entity, social media platforms, or text messages, or through carrying out Computer-Assisted Telephone Interviews (CATI) through a dedicated computer system.
- Various software suits are used in terms of designing data entry programs.
- Upon designing software, the following points shall be taken into account:
 - 1.The software shall include all the poll questions upon making sure that no question has been forgotten.
 - 2.The data entry screen shall be as similar as possible to the questionnaire pages.
 - 3.The skipping buttons shall be featured under all the questions for which answers do not apply as well as for all the questions which feature transitions.
 - 4.The auditing rules shall be featured under the software, whereby such rules shall feature the following:
 - Range check rules for each question;
 - Consistency and Comprehensiveness check rules: being verifying that all data is inputted in a complete manner, that no question is missing, that the answers are double checked for logicity, example: links are established among the age and educational qualification of the respondent, given the auditing program does not allow the entry of the data pertaining to any individual whose age is ten years old and hold a doctoral degree. Usually, the software would alert users in case any data was inputted in the wrong manner.
 - Additional tools will be elaborated for auditing the inputted data, whereby such software is developed upon completing the collection of all the data, whereby this type of auditing process is called "Batch Editing," which would ensure meeting a higher accuracy and quality level for the inputted data.

5. Designing the Training Program and Elaborating Booklets

During this stage, the training program is designed for all the persons involved in the Poll, including enumerators, supervisors, and observers, whereby such program is designed in case the methodology adopted for data collection relates to the person or phone interview. In terms of designing the program, the focus shall be attached to the following:

- Highlighting the importance and objectives of the survey;
- Adopting various educational methods aimed at developing the skills of the research team in terms of conducting a personal interview with the respondents;
- Providing training on asking questions to respondents through role-playing sessions, whereby one of the enumerators shall assume the role of a respondent while the other would ask questions;
- Introducing all the teams to their roles, tasks, and responsibilities, as well as to the work mechanisms.

6. Forming the Data Collection Teams

This stage is implemented in case the methodology adopted for data collection relates to the person or phone interview.

- The teams are formed in a manner that would contribute to increasing the efficiency of the data collection process in an optimal way, whereby it is important to ensure achieving alignment between the number of participants to the data collection process and the timeline set for completing the fieldwork.
- The geographic distribution of the required sample and the closeness or distance shall be considered in this respect, whereby the field teams shall be formed in a manner that would ensure that each team is competent within a specific area or more.
- The gender of the enumerator and the type of the Poll shall be taken into account. In fact, some households do not accept male individuals are present while involved in data collection. In this case, data is collected by female enumerators.

In light of the aforementioned, the number of enumerators and teams required for any poll is calculated.

7. Sample Selection

The sample is selected in accordance with the sample design, upon relying on the identification of individuals who form the sample frame.

8. Carrying out the Pre-Test for the Data Collection Tool

This test aims at verifying that the poll tools and other tools are being correctly used without facing any problem or failure and that, in terms of the accuracy, consistency, and clarity of the questions, the extent of the response to answering the questions by the respondents, and the estimation of the period required by the enumerator to complete the questionnaire. Moreover, this stage also encompasses the identification of open-ended questions to transform them into closed-ended questions as well as all the problems which may be faced in the field.

9. Reviewing the questionnaire and all Booklets

In light of the results drawn from the pre-test carried out in terms of the Poll and the relevant tools, the strengths and weaknesses of the Poll, which were highlighted during the actual implementation of the process are identified, the Poll is reviewed along with all the instructions booklets if needed, as to ensure that they are drafted in their final form.

2.3. Opinion Polls Implementation Stage

There are a set of procedures that must be implemented and verified prior to drawing the finding of a poll, whereby such procedures encompass following up and monitoring data quality, developing the fieldwork mechanism and a system for communicating between the teams, verifying and analyzing them, and carrying out a number of reviews as to review the Poll at the level of the field, in order to verify accuracy and that no individuals were neglected. This stage is followed by carrying out the desk review process in terms of this data, whereby some interviews are repeated in order to make sure that the enumerators have fulfilled their duties.

1. Following up and Monitoring quality

Data shall be considered as one of the inputs pertaining to the decision-making process. Hence, meeting the required quality level during the data collection process is essential, as such achievement positively or negatively impacts the decision-making process.

During this stage, maintaining a high level of accuracy and quality shall be taken into account, whereby the relevant costs and efforts to be deployed shall also be considered.

Furthermore, many ethical considerations shall be taken into account upon implementing the Poll, including the following:

- Preserving the safety of the respondent to the Poll and ensuring that the data disclosed under the Poll are not used in a way that would inflict damages upon the participants;
- Informing the participants of the actual objective of the Poll and disclosing the implementing and funding entity in a transparent manner, in case the participants wanted to know its identity;
- Upon announcing the results, the implementing and funding entity of the Poll shall be disclosed;
- Results and findings shall not be used in a selective manner to impact public opinion and ensure that the Poll is a way for measuring public opinion rather than being a means for influence.

2. Data Collection Mechanism

During this stage, data collection teams are formed, which consist of enumerators, supervisors, and coordinators.

3. Data Collection

During this stage and after the data collection teams are formed, data collection is started.

4. Review Response Rates

- Estimating the response rates and ensuring early identification of the reasons for the response rates decrease in terms of previous polls, whereby such decrease may be due to the fear of respondents or their doubt in the grounds for implementing the Poll;
- Identifying and analyzing all the characteristics and properties of the individuals who would refuse to respond as to avoid facing problems in the future; and
- Most importantly, preserving high response rates along with maintaining a high quality and accuracy level in terms of collecting data.

5. 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.

a. 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.

b. 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.

6. Re-Interview of respondents

Usually, 10% of the interviews which were randomly carried out by each enumerator who has collected data are often repeated, whereby the individuals of a sample are visited again in order to verify that the enumerator has abided by the required sample and obtained information from the required person, as well as to verify the answers given to some questions upon comparing them with the original questionnaire. The reviewing process may also be conducted via phone calls in case a phone number was available for the respondent.

This stage is beneficial in terms of measuring the burden imposed by the respondent upon the sample individuals, as well as identifying the most appropriate timings for conducting interviews and the best means of communication with the respondents.

2.4. Processing and Findings

1. Performing the Coding Process

Coding, in this respect, may be defined as assigning a number or code to each answer, whereby this step is undertaken by one of the specialists in terms of reviewing the coding guide, namely when it comes to open-ended questions, as the respondent often provides answers which express his opinion in an open manner. In this case, the encoder would elaborate a detailed guide for all the answers, under which he shall grant a uniform number for all the similar answers.

Moreover, some economic activities and professions for which the enumerator was not able to collect their data would be assigned a code to identify their category, whereby the relevant documents are printed on the field, and they are assigned a code in the offices.

During the coding stage, the use of the national and international coding guides adopted by the statistical entities are referred to as to ensure that comparisons are carried out in terms of the analysis results.

Usually, the coding guide is elaborated while the data entry software is being elaborated for the questionnaire, whether they are being elaborated on iPad devices or websites.

2. Designing Poll Findings

During this stage, each of the data is assigned a specific description, pursuant to a Data Dictionary under which the questions and variable label are defined along with the content related to each value label question.

This stage also relates to deriving new variables which are not drawn from the poll questions, such as transforming age-related data into large age groups or five-year age groups or transforming income or expenditure-related data into groups, among other examples.

3. Opinion Polls Data Analysis

It is known that selecting the appropriate analysis method for the available statistical data depends on the main objective of the analysis process, which governs the depth of the required analysis, such as the cases where the analysis is a descriptive analysis for the general trends of data, which is based upon calculating some indicators which summarize data, or cases where data is presented in the form of charts, whereby the charts would describe the general direction of such data in terms of increases, decreases, or fluctuations, among others, or cases where analysis is an in-depth analysis carried out within the framework of an inductive analysis which attempts to link between the various variables and thus, develop multiple regression models which determine the relationship between variables.

On the other hand, the type of variable in which data is to be analyzed assumes an important role in terms of determining the appropriate analysis method and the type of indicators that may be represented. In fact, quantitative variables are adequate for representing quantitative indicators, whereas qualitative variables may not be applied in terms of quantitative indicators.

This chapter tackles the statistical analysis methods which are adequate for analyzing the data pertaining to opinion polls and focuses on the descriptive statistical analysis involving the relevant indicators, calculating the correlations or linear factors among variables, and carrying out statistical tests, such as the Chi-Square Test.

Types of Statistical Data

Statistical data are divided into various types, which are determined pursuant to the type of variables represented by such data, as there are continuous data and discrete data. Moreover, there also is nominal data related to nominal values that classify values into various types, and other ordinal data which aim at arranging data or the values of variables.

On the other hand, data may also be divided into interval data reflecting the differences or additions in terms of their values which bear statistical connotations, as well as into ratio data which reflect the result of dividing such data pursuant to their connotations.

Discrete Data (Qualitative Data)

Discrete data represents the set of separate and different data which are expressed by integers. In other terms, such data refers to the information which may be counted (1, 2, 3, ...). This type of data includes, for instance, the number of patients who refer to a specific doctor, gender (male, female), and blood type (O, A, B, and AB), and is generally related to the type of variable values rather than their quantity.

Continuous Data (Quantitative Data)

Continuous data represents the set of data in which values represent the set of real and fractional values which may be classified under categories. This type of data includes data related to expenditure, income, age in years and months, and temperature and is related to the quantity of the variable values rather than their quality.

Nominal Data

Nominal data refers to categorical variables, whereby each of the values represents a different level of categorization, upon considering that the categorization value is meaningless, but represent numerical values which enable the classification of cases as per various levels, whereby such cases may not be arranged. For instance, the marital status of individuals may have various codes, whereby code (1) may be assigned to single persons, (2) may be assigned for married individuals, and so forth, and code (0) may be assigned to males and (2) may be assigned to females in terms of genders.

Ordinal Data

Ordinal data refers to the descriptive data related to variables on ordering. They are sometimes called ordinal data, as they may be arranged in ascending or descending order. For instance, all the questions related to the level of satisfaction of the public for a specific service are considered ordinal data. Hence, (5) would mean very satisfied, (4) would mean satisfied, (3) would mean neutral, (2) would be unsatisfied, and (1) would mean very unsatisfied in terms of measuring the satisfaction level.

Interval Data

Interval data refers to data related to a specific variable, whereby the extent of variability among those values or their total would be meaningful, as it would enable distinguishing those values. For instance, if the first student obtained a grade of 90 and the second student obtained a grade of 80, the difference among both grades would be 10, which is a meaningful value. In other terms, the first grade is more than the second grade by 10 points. However, if the relationship to the head of the household were a brother (Code 3), and the code assigned to the head of the household was (1), the difference among both values, being 2, would be meaningless. Hence, these data would not be considered interval data.

Ratio Data

Ratio data refers to data related to a specific variable, whereby the extent of variability among those values or their total as well as the result of their division would be meaningful, as it would enable drawing a distinction among those values. For instance, in case the income of the first person amounted to 20 thousand Dirhams and the income of the second person amount to 40 thousand Dirhams, the ratio of the first person's income to the second person, being the result of dividing the first person's income by the second person's income would amount to 50% and would, hence, be meaningful. Therefore, these data on income would be considered ratio data.

In addition to the aforementioned, it shall be noted that all ratio data meet the interval data properties, but the opposite is not correct.

Methods for Presenting Data

Presenting statistical data is one of the methods for interpreting and clarifying data and reading their results accordingly. Hence, this method may be considered an essential auxiliary method for analyzing statistical data. There are many methods and means for presenting statistical data, the most common thereof being the following:

Frequency Table

Frequency tables represent a method or mean for summarizing a set of data. It is a record featuring the frequency of all the values of the variable, which will be included in a table. It is possible to add the percentage of frequency under the frequency table to produce a proportional frequency period.

Moreover, frequency tables are used for summarizing the values of interval, nominal, and ordinal values, and they may also be used for summarizing the values of continuous data upon dividing them into categories.

However, in the case of data for more than one interval variable, the frequency table would then become a Homogeneity Table, as the variables' values are listed on the rows of this table while its columns would feature another variable.

Example: Assume that the following numbers represent the results of an opinion poll carried out for 25 individuals in terms of a specific phenomenon:

5, 1, 2, 3, 3, 2, 3, 4, 4, 3, 2, 2, 5, 5, 5, 4, 4, 5, 4, 4, 2, 5, 1, 3, 1.

The following frequency level is a summary of the answers given for a question on the assessment of agreeing upon a specific phenomenon through coding, whereby this variable is expressed as follows, according to absolute frequency and percentage frequency for each of the cases:

Status	Frequency	Percentage (%)
Strongly disagree (1)	3	12.0
Strongly Agree (2)	5	20.0
Neutral (3)	5	20.0
Agree (4)	6	24.0
Strongly Agree (5)	6	24.0
Total	25	100.0

However, in the case of continuous data pertaining to the ratio or interval data, a frequency table may be elaborate in this respect, whereby these data may be listed as non-intersecting continuous data.

As for the number of categories which may be divided into raw data groups, it depends on the type of variable represented by such data and their extent, whereby it is known that the extent is the largest of the raw data values after subtracting the smaller value.

Example: The following data represent the ages of individuals in a specific population. These ages may be represented under a frequency table where the values represent age categories rather than absolute age values, as the case is in the following table.

3,2,5,4,3,7,9,8,5,4,10,12,14,14,13,11,10,11,4,5,6,8,4,2,3,1,1,1,7,7,8,9,6,7,8,10,3,10,4,5,6,8,3,2

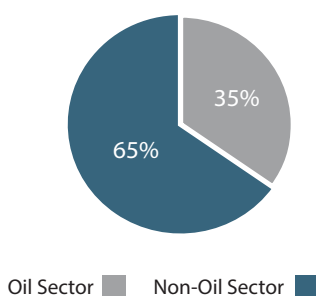
Age Category	Frequency	Percentage (%)
0 - 4	16	36.4
5 - 9	18	40.9
10 - 14	10	22.7
Total	44	100.0

Pie Chart

Pie charts represent a technique or method for summarizing a set of raw interval data. It is represented by a circle divided into the sector, where each sector represents a specific category or stratum, and where the area of each sector (or the angle value of the sector) would be proportional to the number of the value pertaining to the category represented by this sector.

Example: The following pie chart reflects the percentage of contribution to the Gross Domestic Product upon adopting the current prices at the Emirate of Abu Dhabi during the third quarter of 2015. The chart would lead one to notice that the sector representing the GDP of the oil sector covers an area of 35% of the circle. Hence, the angle of this sector would represent 35% of 360, thus 126 degrees, being a wide-angle, as shown by the chart below.

Sector Representing the Gross Domestic Product

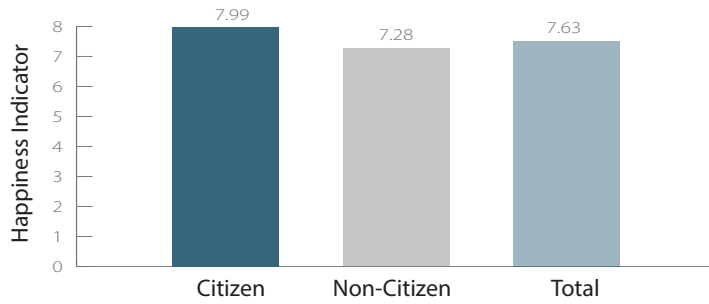


Bar Chart

The bar chart is a technique or method for representing interval or nominal data, and which is popular in terms of statistical processes as it helps in identifying the type and characteristics of a statistical distribution for a set of data, whereby it presents data in the form of a group of bars which have equal width, and where each bar represents a category or stratum of the variables' categories. The height of the bars represents the frequency of the values about the category represented by such a bar.

Example: The following bar chart represents the residents' happiness indicator within a specific population, as per their nationality, whereby the national citizens have achieved a 7.99 level out of ten, while non-residents have obtained 7.28 points, whereas the overall indicator of the population amounts to 7.63.

Happiness Indicator per Nationality

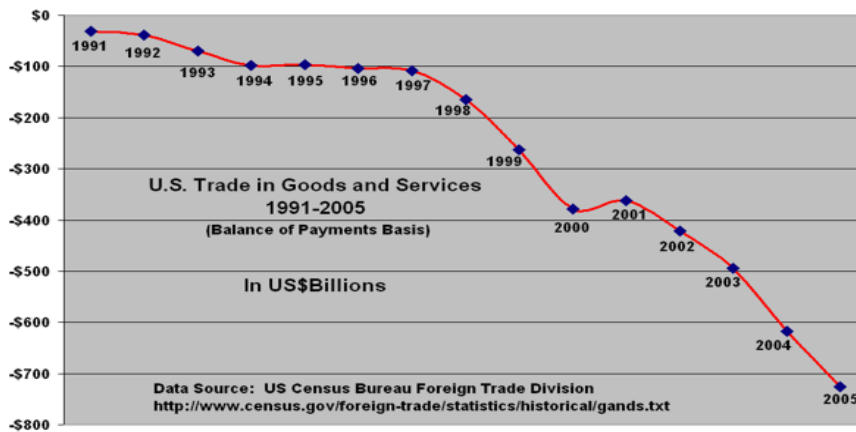


Gradient Descent

Gradient descent are used for representing continuous data related to pair variables, whereby the first variable values would be represented on the horizontal axis and the second variable values would be represented on the vertical axis, where the ordered pair points of both variables are connected through a gradient descent line.

This type of graph is often used in terms of following up on the developments of a specific phenomenon, through following up on the directions of a specific variable in light of the changes undergone by the value of the other variable, as to highlight the relationship between them, which is the case of time changes about the data of a specific variable.

Example: The following gradient descent graph represents the value of trade in goods and services in the United States of America between 1991 and 2005 and its development.

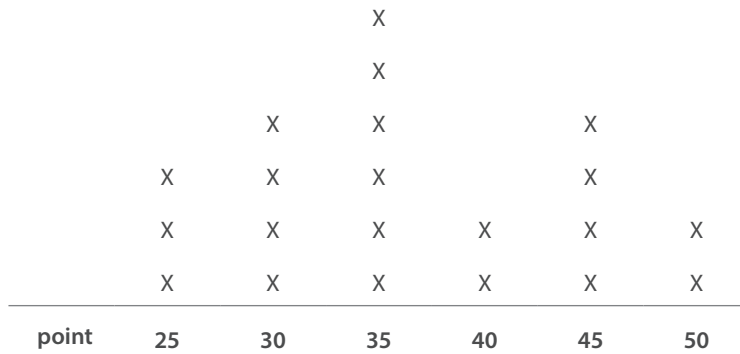


Dot Plot

A dot plot is a technique or method for representing data, whereby it is used in terms of analysis processes to identify the type and characteristics of the statistical distribution of data.

This method is similar to the bar chart method, where dots are shown vertically, according to the frequency of the values represented by the data. This method also assists greatly at identifying the extreme values or gaps within the set of data.

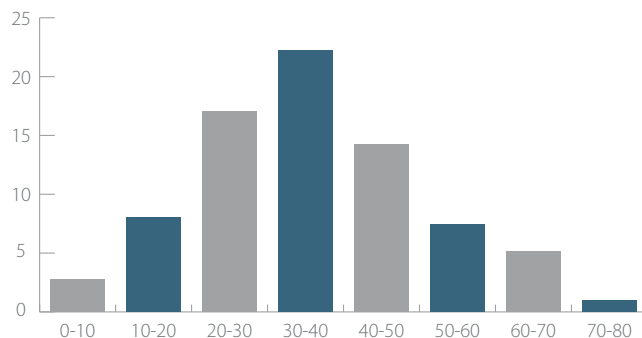
Example: The following graph represents a dot plot for the income of 20 households (Values are set in thousand Dirhams). Accordingly, it is concluded that the most frequent income amounts to 35, whereas the less frequent income amounts to 50. Moreover, the graph also shows the distribution of data, where it begins with setting 25 thousand as a first value which increases upon reaching 35 thousand.



Histogram

Histograms Graph represent a technique or method for representing and summarizing interval data, which is used for identifying the type and characteristic of the probability distribution of data. It is based upon the division of the data extent upon groups, whereby it develops columns for each of them, as for each column to represent a group, where the width of the column would represent the length of the category, and the height of the column would represent the value frequency under such category.

Example: The following graph presents a histogram for the age category of a specific population, where the column height represents the number of individuals in thousands, while the width of the column represents the extent of the age category. For instance, the individuals of the age category ranging between 30 and 40 amount to around 22 thousand in the population.



Stem and Leaf Plot

The stem and leaf plot are a technique or method for presenting and summarizing interval data, which is used for identifying the type and characteristic of the probability distribution of data in a simple and familiar manner.

The stem and leaf plot is similar to the histogram, but its use is more common when the amount of data is relatively small (less than 100). Usually, it is accompanied by a frequency table representing data.

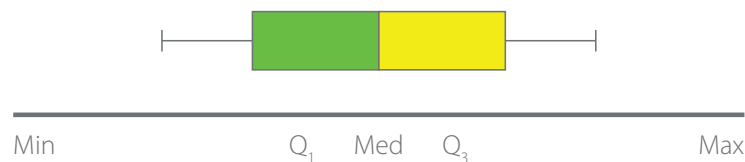
Example: The following plot presents the stem and leaf plot for the grades obtained by 25 students (being the grade out of 100). The first steam represents the number (5), which means that all the grades listed under the case of the tens have five leaves within such steam. Thus, the numbers 55, 55, 56, and 59 are the four numbers that are the least featured under the set of numbers, whereas the numbers such as 100 are among the most frequent grades among the group.

Steam			Leaf			
5	5	5	6	9		
6	2	5	5	6		
7	5	5	8	8	9	9
8	1	5	7	7	9	
9	2	3	5	6		
10	0	0				

Box Plot

This type of representation is a chart representing the distribution and spread of data, which serves to identify the presence of extreme or incoherent data with the main data group. As for developing the plot, it is carried out through determining the first part of data, being the value that encompasses no less than 25% of data, and the third part of data, being the value that encompasses no less than 75% of data and determining the median among them. This step also contains determining the smallest value of the first part and the greatest value of the other part.

In addition to the aforementioned, one may be considered any of the data values falling beyond the box an inconsistent value, which is sometimes called an extreme or deviant value.



Statistical Measurements

The most important statistical methods for summarizing and describing raw data, namely in terms of opinion polls, relate to the elaboration or calculation of easy and expressive statistical criteria, the most important of which include the following:

- **Mean:** Being a sum of set of values divided by the number of views. If the relative importance of views or their weights is not equivalent, the reference weighted average is the product of multiplying each of the values of the views by its weight, then dividing them by the total weights.
- **Median:** Being the value of the average views of the middle data upon ordering them in ascending order. The most important advantage of this indicator is that it is not impacted by extreme or deviant values, as the case is for the arithmetic means.
- **Range:** The range of data refers to the length of the period over which data is spread. Hence, it may be defined as the result of subtracting the greatest value of the data from the smallest value.
- **Variance:** This indicator measures the deviance or difference between the value of views in terms of the arithmetic mean. Hence, it is the average of the square root of squared deviations in terms of values from their arithmetic mean, whereby the square root of this value is expressed by the standard deviation and bears the same objective.

Data Dispersion

In the case of a set of data, the distribution of such data in terms of one of the statistical measures, such as the Arithmetic Mean, Median, or Mode, would be expressed by the extent of spread of the data, As for the scale used to find out the size of its dispersion and spread, it is one of the measures of dispersion, such as the variance, the mean deviation, or the interquartile range.

The spread of the data randomly about the amount of one of the measures indicates that the data is normal, has no high dispersion, and is subject to a normal probability distribution. A normal probability distribution is the distribution or spread of data symmetrically about a mean and about its median at the same time. If the data is represented graphically, it is noted that it takes the shape of an inverted bell, which is the symmetrical shape.

According to probability theory and central tendency theory, in very large populations, data is subject to a normal distribution. Accordingly, many analyzes were built and many theories were drawn within this hypothesis.

Example: The average size of a household in a specific population represents a quantitative variable. Hence, the spread of data would be normal distribution, whereby it is centralized around the arithmetic mean of the size of a household in a population. For instance, in case the average size of a household is (6) individuals, the aforementioned would mean that the number of members of each of the population households is distributed around the number (6), whereby some households may include a lower number of members while others may include a larger number.



Data Skewness

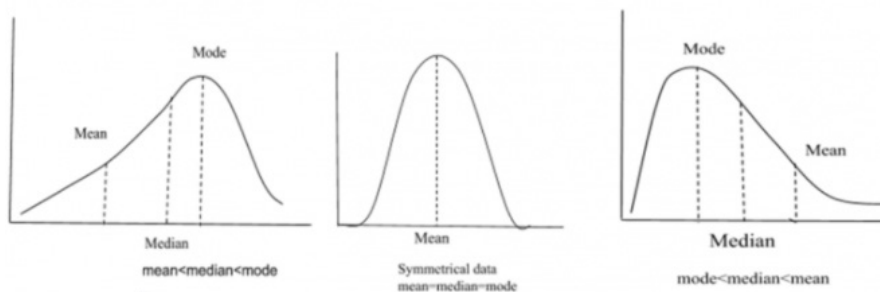
In case the distribution of data was not normal, whereby they are not centered around one of the measures of central tendency, and they are dispersed over a wide range, this case it is called "Skewed data", The magnitude of the skewness is measured by the skew modulus; Where the skew is expressed mathematically by the formula:

$$\frac{(Q3 + Q1)}{(Q3 - Q1)}$$

Whereby Q1, Q2, and Q3 would represent the first, second, and third quarters.

In case the curve skew α was greater than zero, then the data is skewed to the right, whereas in case α was less than zero, then it is skewed to the left, as shown in the figure below:

Example: Most data about the economic variable of a specific population formed of establishments represented skewed data. These include incomes, capitals, intermediate consumption, and other data impacted by the size of establishments, as small establishments often feature variables whose values are less than those of the large establishments.



Inferential Statistics

Inferential statistics are defined as a set of statistical methods aimed at identifying the characteristics of a researched population which is subject to a study, whereby such identification is carried out in terms of a random sample upon adopting specific statistical methods, the most important of which includes making assumptions about the aforementioned probability sample. It is known that Inferential statistics mainly relate to generalizations, estimation, and prediction. However, in some cases, these methods are known for inaccuracy, as they are essentially based upon assumptions. Inferential statistics methods include statistical estimates and testing assumptions, where estimation reflects the value of the arithmetic; meanwhile, testing assumptions refer to approving or rejecting a specific assumption or expression drawn in terms of one of the properties of the population being subject to a study. (To obtain further information in this respect, kindly refer to the Statistical Analysis Principles Guide, available on the Statistics Centre – Abu Dhabi website.)

Inferential Statistics Methods

Statistical Relationships: These identify the nature of the correlation or regression among variables. In mathematics, they are called the correlation coefficients or regression coefficients and are used in terms of studies conducted for two variables which are subject to simultaneous change, whereby such study would aim at identifying the type of relationship among those two variables, such as the study of the relationship between the ages of the students and their academic grades, or the relationship between their weight and height. It shall be noted, in this respect, that there are many types of correlation coefficients, whereby the method used for calculating such coefficient is provided under the Statistical Analysis Principles Guide of the Statistics Centre – Abu Dhabi, or may be applied through any of the ready-to-use software, such as Excel, and SPSS, among others, where it would be easy to apply this method for calculating this coefficient.

Pearson Correlation Coefficient: This coefficient is used to measure the change that occurs in the dependent variable, and is symbolized by the symbol r when the values of the independent variable X change. This coefficient is used in the case of large samples and quantitative data, and it has the following characteristics:

- Its value is zero when the two variables are completely independent.
- The value of this coefficient is positive when the change is direct. Whenever its value is closer to one, the correlation among both variables would be stronger.
- The value of this coefficient would be negative whenever the change among both variables is inverted. Whenever its value is close to -1 , the correlation among both variables would be stronger and inverted.
- It has been agreed that, upon analyzing most of the polls' data, a correlation value ranging between (0) and (0.3) would refer to no significant correlation; a correlation value between (0.3) and (0.5) would show that correlation is weak, a correlation value ranging between (0.5) and (0.7) would reflect an average correlation, a correlation value ranging between (0.7) and (0.9) would reflect a strong correlation, and a correlation value between (0.9) and (1) would show a very strong correlation.

Spearman Correlation Coefficient: This coefficient measures the strength of correlation among two variables represented in the form of descriptive data and small samples which may be ordered, such as the opinions of respondents provided in terms of a specific poll and ordered according to the Likert Scale, where (1) means strongly disagree, (2) disagree, (3) neutral, (4) agree, and (5) strongly agree.

Coefficient of Association: This coefficient is used for measuring the relationship between two phenomena, each of which is divided into two parts or categories, such as gender and nationality (citizen, non-citizen), or smoker/non-smoker, cancerous/non-cancerous, etc., whereby the data are represented under a table formed of four cells. Accordingly, once the Coefficient of Association is closer in values, whether positive or negative, such results would show a strong relationship between both phenomena.

Characteristic	1 st Characteristic	2 nd Characteristic
1 st Characteristic	A	B
2 nd Characteristic	C	D

$$Ass.Coeff = \frac{AD - BC}{AD + BC}$$

For example, 80 individuals were asked this question to establish the possibility of a relationship between smoking and blood pressure. Their answers were as follows:

Characteristic	Hypertension	No hypertension
Smoker	12	28
Non-Smoker	6	34

Pursuant to the aforementioned formula, the Association coefficient would amount to 0.42, which would establish a sort of relationship between smoking and hypertension.

Coefficient of Contingency: This coefficient would measure the strength of the relationship among two phenomena, whereby each phenomenon would feature more than two characteristics, such as examining the relationship between the academic qualification of an individual (Illiterate, primary education, preparatory education, Secondary, etc.) and his income level (high, average, low), where the interrelated table would be formed of more than four cells, upon noting that the compatibility coefficient would range between zero and one and that it would not accept any negative value.

As for the steps for calculating this coefficient, they include the following:

- Square the frequencies in each cell of the table.
- Dividing the square of the frequencies of each cell by the product of multiplying the total of the row by the total of the column including the cell;
- Repeating the previous step for each of the cells, whereby the total is symbolized by C.

In accordance with the aforementioned, the formula for calculating the compatibility coefficient would be as follows:

$$Cont.Coeff = \sqrt{\frac{C - 1}{C}}$$

Example: To establish the relationship between the level of income and the level of expenditure, a study encompassing 100 households was undertaken, and its results were as follows:

Characteristics	High Income	Average Income	Low Income	Total
High Expenditure	10	12	3	25
Average Expenditure	15	20	5	40
Low Expenditure	7	12	16	35
Total	32	44	24	100

Upon dividing the square of the frequency of each cell by the product of multiplying its vertical total by its horizontal total, the results would be as follows:

Characteristics	High Income	Average Income	Low Income	Total
High Expenditure	0.125	0.131	0.015	0.271
Average Expenditure	0.176	0.227	0.026	0.429
Low Expenditure	0.044	0.094	0.305	0.443
Total	0.345	0.452	0.346	1.143

Pursuant to the data listed under the table above, the Contingency coefficient would be calculated as follows:

$$\text{Cont. coeff} = \sqrt{\frac{1.143 - 1}{1.143}}$$

= 0.354

Simple Regression

Simple Regression is a method used for studying the relationship among two variables, the first of which would be dependent variable which value depends on another variable, known as an independent variable. Upon relying on previous data pertaining to the relationship among both variables, the linear relationship, if any, between them is estimated along with the regression line of both. The equation pertaining to this line would estimate the curve of the line β and its intersection with the y-axis:

$$\hat{y} = \hat{\alpha} + \hat{\beta}x$$

Whereby \hat{y} refers to the estimated value of the dependent variable pursuant to the specific value of the independent variable X.

Hence, one of the main uses of the above regression line equation is to make a future prediction of the expected dependent variable values, when assuming the values of the independent variable X according to the above equation.

On the other hand, another criterion is applied for estimating the level of accuracy and efficiency in terms of the interpretation of the independent variable when it comes to interpreting the difference at the level of the dependent variable, whereby this criterion refers to the coefficient of determination R², which is a mathematical square root of the compatibility coefficient between the dependent variable and the independent variable.

In order to obtain further details on the calculation of coefficients and apply their relevant equations, kindly refer to the Statistical Analysis Principles Guide, available on the Statistics Centre – Abu Dhabi website. Moreover, you may also refer to the ready-to-use statistical software, such as SPSS, to elaborate the linear regression line equation upon entering the values of the dependent and independent variables.

4. Presenting Data and Drafting the Relevant Report

Upon verifying that the data is error-free, and upon completing the drawing of all the frequency and crosstabulation tables, the process of presenting data shall begin given the following is taken into account:

- The presentation shall include a summary of the adopted methodology, namely an overview of the sample and survey tools design and testing, as well as a listing of the most important achieved findings, "Summary of the Findings."
- Findings shall be presented in an objective and unbiased manner upon being keen that negative findings are not presented upon highlighting the positive findings only.
- The tables shall be included under Appendices.
- The report shall be supported with simple tables and graphs which clarify the most important finding.
- Comparisons shall be carried out with the results of other polls or polls that were conducted during the previous years, given that the polls were carried out under the same circumstances and conditions.
- All the terms and definitions used in terms of the Poll shall be documented.

3 | Opinion Polls Sampling Methods

The implementation of opinion polls depends on collecting data from individuals in the surveyed community. Therefore, its implementation depends on the provision of a sample of the studied population. This chapter includes a quick review of the concept of samples in opinion polls as an integral part of surveys or statistical work.

More comprehensive details are available in the guide of statistical sampling produced by the Statistics Centre – Abu Dhabi, which is available on Statistical Methodologies page, in SCAD Website.

1. 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:** It is the collection of all statistical units from which the opinion poll 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 Techniques:** It 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.
- **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..
- **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 census or 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 opinion poll 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 “sampling unit non-response,” whereas the “partial non-response” occurs when the respondent only declines to provide answers to specific questions.
- **Standard Error:** It 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 called 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:** it is a method of calculating 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.

2. Sampling Methods

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

2.1. Probability Samples

These samples are selected 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 depends upon 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:

- **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 of the sampling 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 sampling 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 study's population.

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 selecting each sampling unit of the population is given by dividing 1,800 by 60,500, which is equal to 0.02975.

- **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 sample point is reached.

- **Stratified Sample:** A stratified sample is used in terms of ensuring that various strata of the population are represented under the sample. The main idea of a stratified sample is that the information on the study's population is applied to divide such populations into strata with common properties. For instance, let us assume that the population of a study 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 100 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 of a study into many groups, 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.

- **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.

2.2. 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. It is a less stringent method.

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 application of random 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:

- **Convenience Sampling:** This type includes different ways of sample selection, such as selecting individuals who are coincidentally 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.

- **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.
- **Typical Sampling:** Under this type of technique, the most recurrent cases or those featuring a pattern are selected. In terms of many non-official public opinion polls, individuals with a pattern or individuals featuring average properties related to the social and economic variable of a population are selected. However, there are many inconveniences of adopting such sampling techniques. First, how are we supposed to identify the pattern? For instance, in the case of the public opinion polls the individual with a pattern would be a person of average age, educational qualification, and income. Nonetheless, it is unclear whether applying such averages is the most accurate manner of selection as other more important factors, such as religion and race, may also have an impact.
- **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.
- **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,
- **Experts Sampling:** Experts sampling refers to drawing a sample from individuals who are experts in a specific field. Sometimes, this type of sampling is dubbed as the "Committee of Experts." Actually, there are two reasons for experts sampling, whereby this type of sampling may be the best method for obtaining the opinions of individuals who enjoy a specific experience, and the experts sampling would be carried out as to provide more credibility to another sampling method. Accordingly, let us assume that a pattern sampling was carried out, whereby we already know that the criteria and standards for drawing a sample may be subject to some criticism. In this case, the Committee of Experts may be formed of individuals who are knowledgeable of the study fields, whereby they are asked to comment on the sample and its credibility. Hence, the advantage of this method would be the support it provides in terms of defending decisions. However, this method also features inconveniences, as experts are also subject to errors.

4 | Opinion Polls Data Quality

Introduction

Opinion polls represent a type of survey to ensure the fast and immediate identification of public opinion in terms of a specific issue or issue. Moreover, an opinion poll may be defined as a set of questions asked on a selected group of individuals, based upon scientific principles, including the selection of individuals or the formulation of questions. A poll is designed through setting a series of open-ended questions, thus granting the respondent the freedom to answer, or closed-ended questions, being questions providing the respondent with a set of supporting, opposing, or neutral answers, all aimed at identifying the stance of the public in terms of an issue. The aforementioned is carried out by performing a final calculation of these answers to know the number of supporting, opposing, and neutral individuals, thus identifying the public opinion through the selected sample.

Data Quality's in Terms of Opinion Polls

Data quality refers to the criteria for preferring statistical data over other statistical products and services.

Accordingly, data quality in terms of opinion polls would refer to all the fields and aspects related to the extent such data meet the needs of the user and responds to his expectations in terms of performance, level, form, and content, as well as to the high level of information accuracy, in all stages, whereby such accuracy would be aligned with the requirements of data users and decision-makers.

Opinion Polls Statistical Quality's Procedures and Criteria

Preparing quality criteria and procedures about implementing opinion polls shall consider the following:

- Ensuring consistency with the quality guides and criteria of the Establishment, namely the criteria for implementing a statistical survey;
- Preserving the characteristics and features of the poll, such as its simplicity, its fast analysis, results, and costs;
- Considering the nature of the questions asked under the opinion poll.

Furthermore, the implementation of the statistical quality assurance criteria and procedures shall be considered in terms of opinion polls, as per each of the main stages which will be focused on under this guide in detail, whereby such stages include the following:

1. Planning and Preparation Stage;
2. Data Collection Stage;
3. Data Preparation, Processing, and Analysis Stage.

First: Opinion Poll Planning and Preparation Stage

It is essential to determine the poll objectives and needs for its implementation, as the quality of any process lies within the importance of its implementation. Otherwise, carrying out such an opinion poll would not be beneficial. It is also essential to consider all the available options for obtaining the product of this poll to identify the most appropriate method for implementing such a task. Accordingly, quality-related procedures about this stage would include the following:

- During this stage, major importance is attached to the opinion of opinion polls users to ensure permanent communication and contact with data users, the identification of their opinions, and the extent of meeting their needs.
- Polls' planners shall sufficiently be aware of the previous experiences carried out in this respect, whether at the internal or external level, to benefit from such experiences and avoid incurring any error or face the same obstacles that were faced therein.
- Coordination with the competent entities would be a must, along with carrying out the necessary discussion in order to identify their needs in terms of the poll, namely in the cases where new opinion polls are conducted.

- A highly experienced team shall be formed and involve the participation of the individuals who have taken part in previous polls to ensure that more than one opinion is available in terms of setting the ideal circumstances for work.
- The roles and responsibilities of the team shall be identified along with the adequate number of persons in charge of the survey implementation to ensure the achievement of the latter within the set deadlines. The needs of the users shall be aligned with the incurred costs in terms of meeting such needs and the response burden.
- The poll objectives shall be set in line with the needs of the key users and strategic partners, whereby such objectives shall be clearly formulated under a written document, along with the main and detailed objectives of the poll.
- The scope of the poll targeted population and geographic coverage shall be accurately identified in line with the poll objectives.
- The main variables which will be included under the poll questionnaire shall be identified.
- The poll manager shall be sufficiently aware of the stages pertaining to the poll implementation.
- A plan shall be elaborated for implementing the poll, whereby such a plan shall consider the circumstances under which the poll would be carried out.
- An accurate study shall be carried out in terms of the main elements upon which the poll is based, such as the supporting equipment, including computer devices, landlines, and Internet, among others.
- The questionnaire design is considered one of the key stages of implementing a poll. Accordingly, it would be essential to set specific principles for the elements and features through identifying the concepts, content, and features that the questionnaire will cover.
- Hence, a verification process shall be carried out to ensure that the questions meet all the objectives set for the poll and that there are no other questions unrelated to the poll's goals and indicators.
- In addition to considering the elaboration of the poll questions clearly and simply, the selection of suitable and easy terms, and the writing of the poll question logically and consistently, as to ensure obtaining correct and accurate answers, the length of the questionnaire and the number of questions shall also be taken into account, as the aforementioned would impact the quality of collected data and the response rate.
- Furthermore, the questionnaire shall be consistent in a way that would ensure that the errors of the respondent are immediately spotted during the data collection process, whereby the questionnaire would include questions which would help at identifying errors or auditing information and that, namely in terms of the studies or surveys related to collecting data about topics or previous events, the thing that would require the respondent to reiterate providing a correct answer, or in the cases involving collecting data from seniors.
- The questionnaire shall not include questions biased toward a specific answer, nor shall it feature any bias in terms of asking questions by the data collector.
- The questionnaire shall be divided into sections as per the tackled sub-topic.
- The questionnaire shall bear a title, an introduction of the Centre, a mention of the main objective of the poll, a pledge guaranteeing the protection of the respondent's privacy and the preservation of information confidentiality, whereby such pledge shall be featured at the beginning of the questionnaire as to invite respondents to interact in terms of the poll.
- A code, special number, or serial number shall be assigned to the questionnaire to facilitate the follow-up process.
- The questionnaire shall be designed in line with the data collection process (electronic questionnaire, paper questionnaire, self-filling questionnaire, etc.). (Kindly select the type of interview and questionnaire)
- The questionnaire shall be designed in a manner that would ensure it is self-coding (whereby the codes would be available in the questionnaire during the data entry process) whenever possible. If questions require numerical answers, the measurement units and multiplication coefficient shall be mentioned whenever necessary.

- The name of the enumerator responsible for filling the questionnaire and the other supervisory teams who have reviewed, audited, and adopted the form shall be mentioned.
- The key instructions shall be attached to the questionnaire itself, whenever appropriate, given that such instructions are clear, easy, and easily accessible, whereby these instructions would ensure the data collection process' accuracy and consistency.
- The use of the questions and reference periods shall be verified, along with ensuring that asking ambiguous questions is avoided.
- The guides and criteria adopted at the Centre shall be used whenever appropriate.
- A system shall be elaborated for field review and auditing.
- The work methodologies pertaining to previous similar polls shall be referred to as to draw upon previous experience if possible.
- A specific mechanism and methodology shall be set for the data auditing process carried out throughout the data collection, processing, and analysis processes to verify the inputted data's accuracy.
- Modern techniques and technologies shall be used in terms of data entry, whenever appropriate, given that such tools are thoroughly and sufficiently tested prior to carrying out fieldwork, as to avoid any failure thereof while performing data entry, as such failure would expose data to the risk of featuring errors.
- Data entry software shall be provided with automated compatibility features to alert the data entry clerk upon entering illogical or erroneous values to avoid errors in terms of data entry.
- An appropriate mechanism shall be set for enabling access to the methodologies and work guides used in terms of polls for all the competent individuals in this respect.
- The design of the sample is among the most important priorities about data quality, whereby sampling surveys are expected to provide a visualization of the population to be examined as a whole. Accordingly, there are three main questions to be answered upon designing a sample:
 - What is the study population?
 - Who are the study elements (individuals, households, establishments) who are encompassed by this population?
 - Are the frameworks about this population available?
- The framework for selecting the required sample shall be a recent and comprehensive framework that does not include any frequent or missing units, whereby adopting an old, non-updated, and inaccurate framework would negatively impact the quality and accuracy of statistical data. Furthermore, all the elements of a population shall bear an accurate description under the aforementioned framework to ensure easy access to them.
- In terms of areas, maps shall be applied upon the real geographic scope to ensure that geographic zones' borders are clear and aligned within the relevant framework. In this respect, modern techniques may be used in terms of geographic information systems.
- The best sampling design would be establishing a balance between the accuracy of results and the cost of the survey, whereas incurring the lowest costs possible in terms of implementing a survey shall be subject to meeting the required accuracy level and implementing a survey with a minimum amount of errors, namely sampling errors, being errors which would emanate from data which were collected from a sample rather than from a comprehensive survey process. Hence, it is preferable to adopt the probability sampling method whenever possible. Upon determining the size of a sample, the elements impacting the identification of the sample size, such as the targeted, expected, or calculated non-response rate, as drawn from previous polls, the confidence level required from the results, the sampling error desired to be achieved, the impact of design in terms of multiple-stage cluster samples, the level of details in terms of disseminating the results of

a poll, and the extent of difference and variability in terms of the variable's value or other variables, among others.

- The sample design shall consider the targeted, expected, or calculated non-response rate, as drawn from previous surveys, as well as the confidence level required from the results, the sampling error desired to be achieved, and the dissemination levels.
- Adopting a scientific methodology and abiding by the accurate scientific principles would increase the efficiency of the sample and the accuracy of the estimates emanating thereof, whereby calculating the sampling errors featured in terms of previous surveys or studies is beneficial in terms of determining the ideal sampling design and the best sample size.
- The efficient sample design shall be ensured to guarantee that the highest level of accuracy is met upon incurring the lowest cost possible, that the design is characterized by its easiness, simplicity, and applicability, as well as by its flexibility and the possibility of carrying out modifications at the level of the designed.
- Adopting a scientific methodology and abiding by the accurate scientific principles would increase the efficiency of the sample and the accuracy of the estimates emanating thereof, whereby calculating the sampling errors featured in terms of previous surveys or studies is beneficial in terms of determining the ideal sampling design and the best sample size.
- Elaborating indicators for measuring the framework coverage increase or decrease is a must.
- The sample distribution upon the various population segments and stratum shall be considered. The survey would provide a summarized visualization of the study population upon adopting the relevant scientific methods that encompass the probability sample methods that enable the measurement and calculation of errors therein.
- Errors impact data quality and cause the production of biased estimates resulting from omitting or repeating some sampling units. This error is closely linked to the sampling frame. As aforementioned, whenever the sampling frame is comprehensive, recent, and accurate, the used sample would be representative, and the emanating estimates would be unbiased. Moreover, inclusiveness error would lead to producing biased results as a result of the lack of repetition of the sampling units, or the presence of a difference or variation in terms of calculating the estimates required to be drawn from this survey, as some unnecessary individuals were listed under the frame, or errors related to assisting information were listed under the frame that was adopted for drawing a sample, whereby all of the aforementioned would impact data quality.
- The sampling errors findings pertaining to previous similar polls shall be referred to as to improve the sample design of the current poll.
- Comprehensive documentation shall be available in terms of the stages related to the sample design and drawing, whereby such documentation would enable data users to understand the findings of the poll, become aware of the various details, such as the coverage rates, response rates, accuracy levels, and calculate the sample size and its distribution, as well as weighting.

Second: Data Collection Stage

- Once a plan is elaborated for carrying out fieldwork, identifying the environment where the poll would be implemented is a must, whereby such environment would encompass the identification of the natural and geographic characteristics of the targeted regions to consider their properties. Furthermore, the characteristics of the sampling unit, being either individuals, households, or establishments, shall also be identified. The persons in charge of the polls shall be provided with good training in terms of identifying the selected sampling unit and the data collection methods prior to beginning the carrying out of fieldwork, and they shall also be subject to training related to the interview methods and obtaining information in the correct manner.
- There are many means of data collection, including personal interviews or collecting data through phone calls or e-mails. Hence, caution is key, along with training workers to collect data in a good manner, each pursuant to the method adopted in this respect, whereby personal interviews require communication skills to ensure that the respondent's confidence is earned within a climate of relaxation for the respondent.

Furthermore, workers shall also be provided training in terms of the style of asking questions and being unbiased, namely toward a specific answer, which might impact the respondent's answer or cause another problem.

- A plan shall be elaborated for the data collection process to achieve the best data accuracy levels, mitigate the burden imposed upon respondents, decrease costs, ensure abidance by the timeline, and guarantee the availability of an alternative plan to address any potential problem.
- Supervisors and observers shall ensure that the data collection tools, achievement reports, and regulatory reports templates are all available along with all the relevant documents and instruction booklets.
- Work shall be organized, and the tasks and responsibilities shall be identified in a manner that would ensure that workflow is carried out smoothly upon being subject to field follow-up and control as well as to quality control.
- Fieldwork zones and sampling units shall be determined and identified accurately, whether at the level of maps or in the field, to ensure that no sampling unit is ignored and that other visits are carried out.
- The sampling units shall be distributed among enumerators pursuant to the expected achievement rates.
- Enumerators shall be distributed upon supervisors to ensure that the project is achieved within the set deadline and pursuant to the relevant costs, to guarantee that the observers are able to follow up on the work carried out by the enumerators and that the supervisors are able to follow-up on the work carried out by the observers.
- Indicators shall be set in terms of data collection quality control to assess the procedures of the fieldwork.
- Modern and recent technological devices and tools shall be used in terms of the data collection process whenever appropriate.
- Training tools and booklets shall be developed to enable the use of various training methods and the adoption of specific criteria in terms of developing the monitoring skills among supervisors and observers.
- Expertise and qualifications shall be taken into account among work candidates, as well as the assessment results, their level of participation and attendance of training sessions and that, upon selecting the team and determining its responsibilities.
- The nature of the period during which the poll is being carried out shall be considered in terms of its relevant events and circumstance (such as holidays, vacations, and the weather, among others), as well as of any other event which may negatively impact the polls' results.
- The persons in charge of data collection and entry shall sign a confidentiality agreement (data protection document).
- A brochure pertaining to the main instructions for carrying out interviews shall be provided, whereby such brochure shall be short, clear, and easily available for enumerators.
- The visit time shall be scheduled in line with the poll circumstances, whereby information available from previous polls may be referred to, or direct contact may be established with the respondent to schedule an appropriate time for carrying out an interview.
- Upon holding an interview, the identification card shall be clearly displayed, and the enumerator shall introduce himself and the Center, and explain his mission in a summarized and non-exaggerated manner, whereby he shall also reiterate preserving the confidentiality of information which shall not be disclosed, along with other collected personal or individual data or information.
- During the field interview, the person in charge shall have a proper appearance, wear appropriate clothing items which are acceptable in terms of the traditions of a community, gently tap on the door, respect the households he is visiting, and abstain from entering the house unless authorized.
- The enumerator shall avoid expressing surprise upon being provided with an answer by the respondent, and he shall also avoid showing any reaction in terms of these answers.

- Enumerators, supervisors, and observers shall be subject to training in terms of the technical, administrative, and organizational affairs related to the data collection process, whereby such training shall also encompass the poll objectives, work methodology, the mechanism of completing the form and carrying out fieldwork, and the means of using the various poll's tools.
- The enumerator shall implement quick auditing procedures upon the end of the interview to ensure that the respondent has provided all the required data and did not leave any blank fields upon answering the questions asked under the form.
- Periodic and exceptional meetings shall be held with the field team to discuss and review the achievements in terms of implementing the poll, as well as to draft and document the relevant minutes.
- The balance shall be established between fieldwork and the quality of data, whereby the aforementioned shall be beneficial in terms of deciding upon the end of the data collection process or its resuming.
- The field team shall be well aware of the importance of maintaining data confidentiality and that any person who would disclose any data in an unofficial manner, whether individually or through an entity, will be held liable for such action.
- Enumerators shall be subject to training in the most important ethical principles, values, and behaviors to which they shall abide upon carrying out the data collection process.
- The supervisor shall audit the forms in the field, adopt the correct data, and return the questionnaire featuring errors or incomplete ones to the enumerators for addressing such issues in a methodical manner, whereby it is possible to complete some of the missing data through phone calls.
- The discipline of the data collection team shall be monitored in terms of implementing the instructions and guidelines it has received, whether issued under the work guides or through written circulars, among others.
- Achievement and workflow reports shall be drafted and submitted before the fieldwork officials to ensure that all the sampling units were visited and that all their data were collected.
- A verification process shall be carried out to ensure the presence and implementation of the appropriate methodology and procedures for addressing non-response cases to decrease them.
- Response rate, refusal rates, and non-compatibility rates shall be measured and monitored.
- Logical and linguistic errors featured under the form shall be proofread and edited pursuant to a methodology set by a poll team. Regulations for ensuring data security and confidentiality during the data storage and transfer, namely through using modern technologies, such as tablets, handheld devices, and tables, shall be elaborated.
- The presence of procedures that would ensure the protection of data privacy and its protection against being lost is a must.

Third: Data Preparation and Processing Stage

- The data entry system shall be thoroughly examined prior to being used, and the development processes shall be carried out pursuant to the examination results.
- Adequate training shall be provided to the data entry enumerators, and the appropriate data entry documents and guides shall be prepared.
- Electronic software and applications shall be developed for facilitating the auditing and processing operations.
- Appropriate compatibility rules, which are able to detect any lost data, illogical data, erroneous data, or extreme data, among others, shall be applied.
- The entered data shall be reviewed and compared with the data listed under the paper questionnaire to detect the entry-related errors, analyze them at the entry level, and identify the means of improving such data.
- Security procedures shall apply at the level of data upon being electronically inputted to ensure that no one has access to the data files for review and manipulation purposes.

- The presence of documentation pertaining to the data auditing and processing methodology is necessary.
- The appropriate auditing and editing process shall be immediately carried out simultaneously to the data collection and entry processes.
- Sources and types of errors shall be identified, such as measurement errors, coding errors, and input errors, among others, along with the identification of the appropriate modification methods to be applied as per the error type.
- Non-compatibility rates, erroneous values, and extent of variability caused by the modifications shall be identified, the relevant results shall be analyzed, and improvement plans shall be suggested, given the aforementioned is documented under the auditing report.
- Bias caused by the imputation pursuant to values drawn from assumptions that may have been concluded based upon inappropriate models shall be avoided.
- A method shall be applied for estimating indicators in a manner that would be aligned with the sample design and poll methodology.
- A verification process shall be carried out to ensure that statistical data are ready and prepared in terms of processing erroneous values and lost data, as to carry out their relevant analysis process, and as to ensure that the main data is available under the data file.
- A verification process shall be carried out to ensure that data is appropriate to be used in terms of the analysis process being carried out, whereby such data shall be prepared pursuant to the methodology required by such process, and the appropriate analysis software shall be used.
- The appropriate analytical methods shall be used for the data, and the available ready-to-use software shall be used for implementing such methods.
- The consistency of the poll results shall be tested in terms of the data drawn from the polls carried out during previous years, or data drawn from other sources, such as another poll, or data drawn from administrative records.
- A verification process shall be carried out to ensure the consistency of data listed under the same table, as well as the consistency of the data listed under different tables.
- In the case of data dissemination, the measurement units pertaining to the elements forming tables and charts shall clearly be featured to enable their clear understanding and interpretation for users, upon ensuring that such graphs and tables bear clear and understandable titles.
- The previously agreed-upon tables and indicators that were designed pursuant to the poll objectives shall be retrieved.
- The presence of an organized methodology for the management and archiving of the poll data and documents is necessary.
- If necessary, a report shall be drafted to serve all users, whereby other detailed reports, such as summaries or other specialized technical reports, may be made available.
 - Importance shall be attached to sample design, sampling and non-sampling errors, response rates, and coverage.
 - Importance shall be attached to the design of the poll tools.

Data users require information on the quality of data that they will be used to make decisions and draw trusted results. Hence, the dissemination of comprehensive information on data quality is preferable upon publishing data to enable users to provide the necessary interpretations and analyses upon taking the aforementioned information into account.

The time gap separating between completing a survey and disseminating findings shall be decreased, namely, since data often lose their importance and quality in case they were not communicated to decision-makers, policy planners, and data users in a timely manner.

References

Arabic References

1. Othman, Majed, 2011, Measuring Public Opinion: From Theory to Practice, the Arab Republic of Egypt.
2. Al Tahami, Mukhtar et Al., Public Opinion, Arab Republic of Egypt, 2005.
3. Shiha, Najwan et Al., Field Surveys Methodology, Arab Republic of Egypt, 2004.
4. Saad El-Din Jamal. 2007. Public Opinion Polls Applications and Importance in Terms of Decision-Making. Jordan.
5. Statistical Sampling Guide, Statistics Centre – Abu Dhabi, 2015.
6. Statistical Survey Implementation Guide, Statistics Centre – Abu Dhabi, 2016.
7. Statistical Data Quality Guide, Statistics Centre – Abu Dhabi, 2016.
8. Statistical Analysis Principles Guide, Statistics Centre – Abu Dhabi, 2016.

Foreign References

Larossi, Giuseppe. The power of survey design, A User's Guide for Managing Surveys, Interpreting Results and Influencing Respondents, Washington, D.C. 2006.



مركز الإحصاء
STATISTICS CENTRE

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



www.scad.gov.ae

[Twitter](#) [YouTube](#) [LinkedIn](#) [Instagram](#) [adstatistics](#)