LARGE-SCALE SAMPLE SURVEY DATA: AN ALTERNATIVE TO POPULATION DATA FOR SUSTAINABLE DEVELOPMENT IN THE 21ST CENTURY.

Dr. Ahmed Audu

DEPARMENT OF MATHEMATICS
STATISTICS UNIT
USMANU DANFODIYO UNIVERSITY SOKOTO, NIGERIA

1.1 INTRODUCTION

Socio-economic development has been defined as a progress in terms of economic and social factors within a geographic unit. Economic development implies raising the level of prosperity through increased production, distribution and consumption of goods and services. Social development, on the other hand, provides comparative information about citizens such as income, poverty, employment, employment security, education, health, crime and civic participation. Socio-economic development indicators have been conceptualized by several organizations like European Commission (2007), UNECE et al (2014) and OECD (2016) to include happiness, quality of life, well-being, living conditions, life situation, human development, social capital, generalized and political trust, environmental concepts (ecological footprints), concept of sustainable development etc.

1.1 Population and Population Data

Statistically, the population is technically defined as the aggregate or totality of the units that constitute a universe of elements that could be animate or inanimate. it is a discrete system of distinct, identifiable and desirable members structurally linked together into a composite architecture that may or may not be so obvious. Population trends today are characterized by an increasing divergence across countries and regions. Whereas many of the poorer countries continue to be characterized by rapid population growth, others that are more advanced in their demographic transition are experiencing rapid population ageing and even population decline in some cases. Furthermore, the world is witnessing increasingly complex international migration patterns and many countries continue to experience very high rates of urbanization. These population dynamics influence development at the national and sub-national levels, but also at regional and global levels. As such, it is pertinent to always gather relevant and vital population data from which government and other stakeholders can source relevant information needed for the planning and sustainability of socio-economic development of a country. Census is a principal source of population or demographic data globally. Data on population size, age-sex composition, fertility and mortality

levels, migration and spatial distribution, urban-rural composition, the proportion of literate and labour force, among others, are generated during census exercise in a country. These data are needed for effective socio-economic development planning and its sustainability. In statistics, the population is conceptualized as the universe from which some members (sample) could be randomly selected through some scientific procedures (sampling techniques and methods), in lieu of complete enumeration (census), as a representative in order to realize some relevant information that could projected against the entire population. The essence of sampling procedures is to obtain relevant and adequate information required to obtain approximate population parameters needed for policy and decision making at minimum cost and with optimum precision.

1.2 Population Data as a Tool for Generating Socio-Economic Data

A census of population collects information on basic population characteristics including size, age, sex, marital status, household composition, family characteristics, and household size. Information is also collected on economic measures including labor force participation, occupation, place of work, employment-related industry, and educational attributes such as school attendance, educational attainment and literacy.

The information obtained from population data can be used to determining socio-economic factors of a country. For example

i. Population Size and the distribution of tax payers in the population and revenues can be used to ascertain the economic level of a country
ii. Population Size can be used to forecast country's socio-economic needs. e.g electricity, housing, food, infrastructure, security etc.
iii. Population size aids in the estimation of the environmental impact of population growth, use of water, land, forest and other resources.
iv. Population Labour force can be used to determine the employment status of the citizens
v. Population Age and Sex distributions help identify segments of the population that require different types of services.
vi. Information on marital status, household composition, family characteristics, and household size of a population help in determining housing needs, understanding spatial distribution and changes in the socio-economic status of the citizens
vii. Educational Attainment and Literacy provide information on the educational skills of the work force. These measures also help planners select the best strategies to communicate with residents
viii. And generally, population data helps in formulating economic, social and administrative policies and decisions on resources distribution, governance, security, trades, investments, industries and forecast the future needs of the population so as to sustain the socio-economic development attained at the present.

1.3 Socio-Economic Data as A Tool For Sustainable Development

Boström (2012) defined social sustainability as the improvement of living conditions for both current and future generations.

Goniadis (2015) defined economic sustainability as the ability of the economy to support and maintain economic growth, but at the same time, it requires that natural resources be used efficiently.

Report published by Brundtland (1987) defined sustainable development as development that meets the needs of the present without compromising the ability of future generations to meet their needs.

Hence, sustainability of socio-economic development could be understood as the ability to ensure economic growth without undermining humans' interests and to meet their needs without harming nature.

Socio-economic of nations changes due to new technologies, changes in laws, changes in the physical environment, ecological changes etc. and this causes changes in the needs and aspiration of the population. Hence, there is need for frequent gathering of population data for making new policies and decisions to sustain the developments.

Ssebunya et al. (2019) distinguished three elements of socio-economic sustainability:

social wellbeing, economic resilience, and good governance.

Present and future social wellbeing of the citizens heavily depend on country's socio-economic data like population size, population growth, citizens' educational level, labour force, demographical data and other relevant population data. Present and future of a country's economic development and its sustainability are hanged on the socio-economic factors like population distribution of tax payers, labour force, employment-related industry as well as economic, social and administrative policies and decisions on resources distribution, governance, security, trades, investments, industries and forecast the future needs. Good governance is a product of good and committed leadership through the conduct of public affairs and management of public resources in a manner that promotes the rule of law and the realization of human rights. However, actualization and sustainability of good governance heavily depend on policies and interventions informed by available relevant population data.

2. SAMPLE SURVEY

2.1 Sample Survey Versus Complete Census

Broadly speaking, information on population may be collected in two ways. Either every unit in the population is enumerated (called complete enumeration, or census) or enumeration is limited to only a part or sample selected from the population (called sample enumeration or sample survey). A sample survey will usually be less costly than a complete census because the expense of covering all units would be greater than that of covering only a sample fraction. Also, it will take less time to collect and process data from a sample than from a census. But economy is not the only consideration; the most important point is whether the accuracy of the results would be adequate for the end in view. It is a curious fact that the results from a carefully planned and well executed

sample survey are expected to be more accurate (near to the aim of study) than those from a complete enumeration that can be taken. A complete census ordinarily requires a huge and unwieldy organization and therefore many types of errors creep in which cannot be controlled adequately. In a sample survey the volume of work is reduced considerably, and it becomes possible to employ persons of higher caliber, train them suitably, and supervise their work adequately. In a properly designed sample survey, it is also possible to make a valid estimate of the margin of error and hence decide whether the results are sufficiently accurate. A complete census does not reveal by its self the margin of uncertainty to which it is subject. But there is not always a choice of one versus the other.

2.2 The Principal Steps in Planning and Executing of Sample Survey

The principal steps in a survey are as follows

1. Objective of the Survey:

The objectives of the survey must be stated in clear, concrete and concise terms. A lucid statement of the objectives is most helpful. It is not sufficient and helpful to express the objectives in a vague and general form. The statisticians can help the users (sponsors of the survey) to formulate the survey's objectives to the user's satisfactions. The statement of the objectives should include the reason for the survey, questions to be covered, level of accuracy required and the expected results. Failure to state the objectives of the survey in a precise form may result in collecting a mass of unusable information.

2. Population to be Sampled:

The word population is used to denote the aggregate from which the sample is chosen. The population to be covered in a survey must be defined in terms of content, units, extent and time.

For examples, in an expenditure survey, the population may be defined as in terms of all persons above 18 years of age, living in household, in urban areas, and during the second quarter of the year 2021. The population to be sampled (the sampled population) should coincide with the population from which information is sought (the target population).

In practice, because of non-response, non-coverage or administrative convenience, the sampled is more restricted than the target population. For example, in the expenditure survey described above, institutional household like prisons, army barracks, university hostels, may be excluded and only those above 18 years and earning a living in private households are sampled. It should be remembered that conclusions drawn from the sample apply to the sampled population. Judgments about the extent to which these conclusions will also be applicable to the target population depend on the supplementary information about the nature of the differences between the sampled and target population.

3. The Frame:

A frame is a list or map of the population to be covered together with its identification scheme from which the sample is selected. E.g. list of individuals, or households, list of centers for continuing education in a survey of out-of-school children, list of farms in an agricultural survey, list of establishment in an establishment survey. Before selecting the sample, the population must

be divided into parts that are called sampling units documented in sampling frame. These units must cover the whole population, identifiable and non-overlapping. Sometimes, a suitable frame may not be available and need to be constructed. In such situation, the procedure for the construction should be clearly described. Also, available frame may be defective, incomplete, obsolete or full of duplications. The defective frame should be corrected for its defect before it can be used for sample selection. The problem of incomplete frame can be resolved by supplement the list from another existing frame. In the case of duplication, any unit appearing more than once should has one of the duplicates retained and the rest deleted.

Obsolete frame should be updated.

4. Data to be Collected:

Identify all the data (study and auxiliary variables, attributes, socio-economic characteristics, geographical locations, e.t.c.) relevant to the purpose of the survey. No essential data are excluded. An overlong questionnaire should be avoided as it lowers the quality of the answers to important as well as unimportant questions. A practical method is to chalk out an outline of the tables that the survey should produce. This would help in eliminating the collection of irrelevant information and ensure that no essential data are omitted.

5. Degree of Precision Desired

The results of sample surveys are always subject to some uncertainty because only part of the population has been measured and because of errors of measurement. This uncertainty can be reduced by taking larger samples, using superior instrument of measurement and good sample design, if cost and time permit. Consequently, the specification of the degree of precision wanted in the results is an important step. This step is the responsibility of surveys' users, though statisticians can be of help is the surveys' users are unaccustomed to thinking in terms of the amounts of errors that can be tolerated in estimates, consistent with making good decisions.

6. Sampling Design

The choice of the optimum sampling design depends to an extent on the cost of survey, degree of precision and operational convenience. The general principle in the choice of sampling design is to reduce cost for a given precision or to increase the precision for a give cost of the survey while at the same time controlling the non-sampling errors.

7. Selection of the Sample and Sample Size

For each procedure of sample selection to be considered, estimate of sample size can be made from knowledge of the degree of precision desired, relative costs and time involved.

8. Methods of Data Collection Measurement:

The methods used to collect survey data have evolved with the change in technology. From face-to-face surveys, telephonic surveys to now online and email surveys, the world of survey data collection has changed with time. Each survey data collection method has its pros and cons, and every researcher has a preference for gathering accurate information from the target sample. The survey response rates for each of these data collection methods will differ as their reach and impact

are always different. Different ways are chosen according to specific target population characteristics and the intent to examine human nature under various situations.

Types of survey data based on deployment methods:

There are four main survey data collection methods – Telephonic Surveys, Face-to-face Surveys, and Online Surveys.

- *Online Surveys*

Online surveys are the most cost-effective and can reach the maximum number of people in comparison to the other mediums. The performance of these surveys is much more widespread than the other data collection methods. In situations where there is more than one question to be asked to the target sample, certain researchers prefer conducting online surveys over the traditional face-to-face or telephone surveys. Online surveys are effective and therefore require computational logic and branching technologies for exponentially more accurate survey data collection vs any other traditional means of surveying. They are straightforward in their implementation and take a minimum time of the respondents. The investment required for survey data collection using online surveys is also negligible in comparison to the other methods. The results are collected in real-time for researchers to analyze and decide corrective measures. A very good example of an online survey is a hotel chain using an online survey to collect guest satisfaction metrics after a stay or an event at the property. Online surveys are safe and secure to conduct. As there is no in-person interaction or any direct form of communication, they are quite useful in times of global crisis. For instance, many organizations moved to contactless surveys during the pandemic. It helped them ensure that the employees are not experiencing any COVID-19 symptoms before they come to the office.

- *Face-to-face Surveys*

Gaining information from respondents via face-to-face mediums is much more effective than the other mediums because respondents usually tend to trust the surveyors and provide honest and clear feedback about the subject in-hand. Researchers can easily identify whether their respondents are uncomfortable with the asked questions and can be extremely productive in case there are sensitive topics involved in the discussion. This online data collection method demands more cost-investment than in comparison to the other methods. According to the geographic or psychographic segmentation, researchers must be trained to gain accurate information. For example, a job evaluation survey is conducted in person between an HR or a manager with the employee. This method works best face-to-face as the data collection can collect as accurate information as possible.

- *Telephone Surveys*

Telephone surveys require much lesser investment than face-to-face surveys. Depending on the required reach, telephone surveys cost as much or a little more than online surveys. Contacting respondents via the telephonic medium requires less effort and manpower than the face-to-face survey medium. If interviewers are located at the same place, they can cross-check their questions

to ensure error-free questions are asked to the target audience. The main drawback of conducting telephone surveys is that establishing a friendly equation with the respondent becomes challenging due to the bridge of the medium. Respondents are also highly likely to choose to remain anonymous in their feedback over the phone as the reliability associated with the researcher can be questioned. For example, if a retail giant would like to understand purchasing decisions, they can conduct a telephonic, motivation, and buying experience survey to collect data about the entire purchasing experience.

- *Paper Surveys*

The other commonly used survey method is paper surveys. These surveys can be used where laptops, computers, and tablets cannot go, and hence they use the age-old method of data collection; pen and paper. This method helps collect survey data in field research and helps strengthen the number of responses collected and the validity of these responses. A popular example or use case of a paper survey is a fast-food restaurant survey where the fast-food chain would like to collect feedback on the dining experience of its patrons.

Data may be collected through mail questionnaire (postal, email, online survey), personal interview, physical observation or vital registration depending on the type of data to be collected, unit of enquiry, the subject matter and the scale of the survey. The decision on the method to be adopted for data collection should depend on the cost, level of accuracy desired and practicability of each method. For examples, data about a person's state of health may be obtained from the statement he/she makes or from medical records/examination. For a survey among educated people or survey of industrial establishments, mail questionnaire may be appropriate provided that the addresses (postal or emails) of such individuals are available. In a survey where a large proportion of the population is illiterate, direct personal interview is recommended as it minimized non-response. In household survey, personal interview can be used. In agriculture survey where there are many small time and illiterate farmers, both personal interview (e.g quantity of crop harvested) and physical observation (e.g physical measurement of areas under cultivation) can be adopted. But where the farmers are literate, crop areas and crop production can be obtained by direct interview.

9. Organization of the Fieldwork:

The organization of fieldwork and supervision is very essential for successful conduct of any survey. Much importance should be attached to the recruitment and training of good quality filed staff. Field interviewers must receive adequate training on the purpose of the survey, methods of measurement to be employed as well as method of handling non-response. Trained supervisors should be used to monitor the field interviewers on a random basis during the course of the field work to improve the quality of data collection. The supervision could be done through spot-check, post survey checks and field scrutiny of the completed questionnaires. Lastly, the returned questionnaire should be properly checked for omissions and inconsistencies.

10. Pre-Test and Pilot Survey:

Pre-test is a systematic trial of each of all the various features of the main enquiry while pilot survey is a small-scale replica of the main survey- a type of feasibility study usually preceded by a series of pre-tests. Both pre-test and pilot survey is a trial run for the main survey. It is very useful to try out the questionnaires, filed methods and to familiarize the interviewers on the job before the main survey.

Pre-test and pilot survey helps to:

i. Assess the suitability of the frame
ii. Give insight to the study population variability which can be used to determine sample size required to achieve desired precision.
iii. Assess the adequacy of questionnaire
iv. Ascertain possible performance of the interviewers
v. have an idea about non-response rate
vi. ascertain possible cost and duration of the survey
vii. discover unrealizable objectives

2.3 Sampling Error

Sampling errors occur due to a disparity in the representativeness of the respondents. It majorly happens when the researcher does not plan his sample carefully. These sampling errors can be controlled and eliminated by creating a careful sample design, having a large enough sample to reflect the entire population, or using an online sample or survey audiences to collect responses.

Sampling errors can be reduced by:

1. **Increase sample size:** A larger sample size results in a more accurate result because the study gets closer to the actual population size.
2. **Divide the population into groups:** Test groups according to their size in the population instead of a random sample. For example, if people of a specific demographic make up 20% of the population, make sure that your study is made up of this variable.
3. **Know your population:** Study your population and understand its demographic mix. Know what demographics use your product and service and ensure you only target the sample that matters.

2.4 Principle of Sample Survey

The theory of sampling is based on the following important principles

1. Principle of Statistical Regularity: - This principle states that a moderately large number of population units chosen at random from a population are almost sure on the average to possess the characteristics of the population. This is because in dealing with large numbers, the variations in the component parts tend to balance each other and consequently the variation in the aggregate result is likely to be insignificant.
2. Principle of Validity: - This principle states that a sample survey design is valid if the validity tests and estimates about the population parameters can be obtained.

3. Principle of Optimization: - This principle states that a sampling technique should be able to produce optimum results in terms of efficiency and cost of the design with the resources at disposal of the surveyors or survey users.

The samples obtained by the techniques of probability sampling do satisfy the above principles.

3. LARGE-SCALE SAMPLE SURVEY

3.1 Description of Large-Scale Sample Survey

Large-scale sample survey is a type of probability sampling approach which involves complex multistage designs that combines stratification, cluster sampling and unequal probability sampling. This sampling procedure is usually adopted to minimize cost of survey due to limited available resources and ensure safety of the interviewers due to insecurity in face-to-face surveys. it is the best approach when the sampling frame is not available but can be constructed sequentially as needed. In large-scale surveys, data are usually extensive in contents and coverage of the target population and sample sizes are often very large. Usually, these surveys are typically conducted by large government agencies or private survey organizations that have experienced interviewers, well-design and tested survey instruments and the expertise to design efficient sampling plans. As a result of these features, the response rates are typically high and sophisticated adjustments are often incorporated into the data to reduce non-response bias that results from systematic difference between the respondents and non-respondents. To make these surveys more representative of census, data obtained are usually reweighted using poststratification to agree with known population distributions of characteristics such as age, sex and race. These surveys often contain panel (longitudinal) components in which the same sample units are interviewed repeatedly over time to allowed the researchers or survey users to follow individual changes over time.

Large-scale sample survey has been in used by several developed countries like USA, UK, Canada, France, Kingdom of Saudi Arabia etc., leading statistical agencies in the developed and developing world, non-governmental organizations (NGOs) and private sector groups as an alternative approach in generating population data. Several computer software, packages and applications have been developed in advanced countries to aid data collection during the conduct of the surveys, processing and analyses. Some of these applications include

i. Census and Survey Processing System (CSPro)
ii. Demographic Analysis & Population Projection System (DAPPS) Software
iii. Population Analysis System (PAS) Software
iv. Rural Urban Projection (RUP) Software
v. Subnational Projections Toolkit (SPToolkit) Software
vi. Tool for Assessing Statistical Capacity (TASC)
vii. X-13ARIMA-SEATS Seasonal Adjustment Program

3.2 Relevance of Large-Scale Sample Survey to Developing Countries: Nigeria Context

Large-scale sample survey especially on social surveys provides an inexpensive and easily accessible source of information compared to population census, on a wide range of topics including education, health, economics, demography, politics and criminal justice.

From the description and definition of large-scale sample survey, the design satisfies all the three principles of sample design. This implies that the estimates of population parameters obtained from the data generate through the design are highly efficient with insignificant sampling errors.

Nevertheless, in this century, countries of the world especially Africa countries like Nigeria, have suffered economic crises due to COVID-19 and other pandemics, insecurity challenges, internal conflicts, money laundries, etc. These crises have significantly reduced economic strength of the nations and consequently, some of the countries may not afford collection of population data needed for national planning and development to achieve socio-economic development and its sustainability.

From the above justifications, information from large-scale sample survey is the best alternative to population data required by policies and decision makers for socio-economic development and its sustainability.

4. CONCLUSION AND RECOMMENDATION

The socio-economic development of a nation heavily depends on policies and interventions informed by available relevant population data. Sourcing this vital information from the population requires manpower, resources, access to population units, technology, etc. However, in developing countries like Nigeria, these requirements are mostly unobtainable due to limited resources, lack of access to technology, high level of illiteracy and other emerging factors like insecurity, outbreak of pandemics, communal crises etc. which can hinder the process of population data collection thereby making collected information incomplete, inadequate, inaccurate and full of errors. Thus, making the realization of relevant and adequate data needed for the actualization of socio-economic development a myth.

Therefore, it is pertinent to seek urgent durable, and efficient alternatives to population data. In this regard, large-scale sample surveys can be a perfect substitute that can provide relevant and sufficient information that will inform policymakers on their decisions for sustainable socio-economic development.

REFERENCES

Ayinde, K. (2020). Statistics, National Planning and National Development. m *Proceeding of Royal Statistics Society Nigeria Local Group,* 26-33.

Bamiduro, T. A. (2020). Population Data for Informed National Planning and Development. *Proceeding of Royal Statistics Society Nigeria Local Group,* 1-25.

Boström, M. A (2012). missing pillar? Challenges in theorizing and practicing social sustainability: Introduction to the special issue. *Sustain. Sci. Pract. Policy*, 8, 3–14.

Goniadis. (2015) Introduction to Sustainable Development; International Hellenic University: Nea Moudania, Greece.

Gupta, S. C. and Kapoor, V. K. (2018). Fundamentals of Applied Statistics. Sultan Chand and Sons Educational Publisher, New Delhi.

Schroeder, T. (2018). Giving Meaning to the Concept of Sustainability in Architectural Design Practices: Setting Out the Analytical Framework of Translation. *Sustainability*, 10, 1710.

Singh, S. (2003). *Advanced Sampling Theory with Applications*. Springer-Science+Business Media, B.V, USA.

Ssebunya, B.R.; Schader, C.; Baumgart, L.; Landert, J.; Altenbuchner, C.; Schmid, E.; Stolze, M. (2019). Sustainability Performance of Certified and Non-certified Smallholder Coffee Farms in Uganda. *Ecol. Econ*., 156, 35–47.