

SOCIAL SURVEY METHODS AND THE DESIGN OF SURVEY INSTRUMENTS FOR SOCIAL AND ENVIRONMENTAL SCIENCES

By

Owolabi, Babatunde Oluwaseyi
Department of Urban and Regional Planning,
University of Ibadan, Nigeria
Email: babatundeoluwaseyi@yahoo.com
Telephone: +2348056362930

1.0 INTRODUCTION AND DEFINITION OF TERMS

In order for us to get a clear understanding of the topic of this paper it is necessary to explain or define the key words therein; these are social, survey, methods, design and instrument. It is obvious however that the most important word in the topic is ‘survey’, and greater attention shall be focused on it.

1.1 SOCIAL

The term "social" is used in many different senses. In one sense, it refers to attitudes, orientations or behaviours which take the interests, intentions or needs of other people into account. In another sense, it means the interactive systems in communities of animal or insect populations, or any living organisms. It also means membership of a group of people or inclusion or belonging to a community of people. Never the less, underlying the different senses of its use, is the fact that it refers to human society, its organisation and interactions.

1.2 METHOD

Method refers to a means or manner of procedure, especially regular and systematic steps or way of accomplishing a task or attaining a goal. Method implies a detailed logically ordered plan for achieving a particular task. The methodical approach is applicable to and employed in all disciplines. Social science inquiry employs the scientific method. This refers to the body of techniques for investigating phenomena, acquiring new knowledge, or correcting and integrating previous knowledge. It is based on gathering observable, empirical and measurable evidence subject to specific principles of reasoning. A scientific method consists of the collection of data through observation and experimentation, and the formulation and testing of hypotheses.

Although procedures for inquiry may vary from one field of knowledge to another, however what distinguishes scientific inquiry from others is that scientific inquiry propose hypotheses to explain observed patterns or occurrences. It goes further to design experimental studies to test these hypotheses. These steps must be repeatable in order to predict dependably any future results. The process must be objective so as to reduce element of bias in the interpretation of the results that are obtained. Scientific inquiry is also characterised by painstaking keeping of records and making data and methodology of a particular test or experiment available to other researchers in order to allow verification and validity of the results earlier obtained and the inferences derived from them. This

procedure often leads to the formulation of new hypotheses. This practice which, Williams (1994) calls 'full disclosure', also allows statistical measures of the reliability of these data to be established.

A hypothesis is a statement which suggests an explanation of an observed event. It is a reasoned proposal of a possible correlation between or among a set of events. A hypothesis therefore is not a statement of fact since it has not been tested to determine its validity. A hypothesis is a fishing expedition in the dark. It is thus not borne out of empirical investigation.

Depending on the field of study or phenomenon of investigation, a hypothesis can be formulated either as a mathematical model or as an existential statement. An existential statement usually suggests that some particular instance of the event being investigated has some peculiarities and causal explanations, which can be generalised or is applicable to the wider world.

The utility of a hypothesis lies essentially not only in its potential to displace existing notions or ideas but also in its ability "to stimulate the research that will illuminate ... bald suppositions and areas of vagueness".

The cycle of statement of initial hypothesis, testing, obtaining result and deriving inferences, verification and formulation of new hypothesis will resemble the following cycle described below,

- i. Define the question unambiguously
- ii. Formulate clear and testable hypothesis
- iii. Perform experiment and collect data
- iv. Analyse the data and obtain result
- v. Interpret result and draw conclusions (which may serve as starting point for new hypothesis)
- vi. Document the procedure and publish the result
- vii. Verify or retest the entire procedure (usually engaged in by other researchers)

1.3 SURVEY

The word "survey" was derived from the Latin words *sur* meaning 'over' and *videre* meaning 'to see' or simply oversee. Survey involves the systematic collection of data about populations especially large or widely dispersed groups of people. It could mean a close or systematic study, inspection or investigation or, a general or comprehensive view or treatment of anything. It involves analysis, examination, inspection, investigation and, review.

Survey means different things to different disciplines and therefore differs in terms of purpose, subject matter, coverage, and source of information. In the field of epidemiology, surveys have been used to study among others, the history of the health of populations, diagnosis of community health, the working of health services, search for the cause of health and disease. In the behavioural or social sciences, survey is used most commonly to

gather information about the beliefs, attitudes, and values of individuals. In land management survey

Survey can be classified broadly into descriptive survey and analytical survey. In descriptive survey the objective is to obtain certain information about large groups. In analytical survey, comparisons are made between different subgroups of the population in order to discover whether differences exist among them that may enable researchers to form or test hypotheses about the forces at work in the population.

Survey may be general such as in census surveys or limited such as in sample surveys. Surveys take many forms, but often use questionnaires and interviews.

1.3.1. Census

A census is the process of obtaining information about every member of a population of interest (not necessarily a human population; e.g. housing censuses and traffic). However the term is mostly used in connection with the systematic and complete enumeration of all the people of a nation or a registration region, usually on a specific date. It is a method used for accumulating statistical data, for research, business marketing and, planning purposes A census is an

1.3.2. Sampling

Sampling is the process whereby information is obtained from the analysis of selected but significant representative percentage parts of an entity (population) chosen randomly, with the aim of making general statements that apply to the entity as a whole, or an identifiable part of it. This is sometimes necessary since due to practical and logistic considerations it may be difficult or out rightly impossible to, for instance interview every person in a State or country.

The sampling process consists of 7 simple stages;

- Definition of population of concern
- Specification of a sampling frame, a set of items or events that it is possible to measure
- Specification of sampling method for selecting items or events from the frame
- Determination of the sample size
- Implementation of the sampling plan
- Sampling and data collection
- Review of sampling process

Sampling has among several advantages considerable savings in time and money. It also allows more resources to be devoted to a much smaller population and thereby generate information that is often more accurate “as greater effort can be expended on the training of interviewers, more sophisticated and expensive measurement devices can be used, repeated measurements can be taken, and more detailed questions can be posed”.

The process of drawing conclusions about the larger entity based on the information contained in a sample is known as statistical inference.

In surveys, various terminologies are used to describe the component parts of the survey process. These include "target population", "sampling unit" and, "sampling frame".

1.3.2.1. Target population is commonly used to refer to the group of people or entities (the "universe") to which the findings of the sample are to be generalised. More often than not we are interested in understanding process or phenomenon with reference to a population and seek to intervene. However and as we had earlier observed, the impracticability of studying the entire population makes it necessary that we select representative samples of the population. It goes without saying therefore that our samples will be influenced to a very large extent by our subject population. Sometimes also, we seek knowledge about the cause system of which the population is an outcome; for example, when a researcher performs an experiment on rats with the intention of gaining insights into biochemistry that can be applied for the benefit of humans. There is thus a need to carefully and precisely define the population of interest. Time spent in making the population of concern precise is often well spent, because it raises many issues, ambiguities and questions that would otherwise have been overlooked at this stage and which if not addressed will affect the result and the inferences derived from the study.

1.3.2.2. Sampling Unit is the basic unit (e.g., person, household,) around which a sampling procedure is planned.

1.3.2.3. The Sampling Frame is any list of all the sampling units in the target population. Alphabetical listings of residents in a community, telephone directory or voters register are examples of sampling frames. The sampling frame must be representative of the population and this is a question outside the scope of statistical theory demanding the judgment of experts in the particular subject matter being studied. All the above frames will definitely omit some people who will vote at the next election and contain some people who will not. People not in the frame have no prospect of being sampled. Statistical theory tells us about the uncertainties in extrapolating from a sample to the frame. In extrapolating from frame to population its role is motivational and suggestive. In defining the frame, practical, economic, ethical and technical issues need to be addressed. The need to obtain timely results may prevent extending the frame far into the future. The difficulties can be extreme when the population and frame are disjoint. This is a particular problem in forecasting where inferences about the future are made from historical data.

There is however, a strong division of views about the acceptability of representative sampling across different domains of study. To the philosopher, representative sampling procedure has no justification whatsoever because it is not how truth is pursued in philosophy. 'To the scientist, however, representative sampling is the only justified procedure for choosing individual objects for use as the basis of generalisation, and is therefore usually the only acceptable basis for ascertaining truth' (Marino 1994?).

1.3.2.4. The Sample size is the amount of sample large enough to be representative of the target population. It is derived from the sampling frame. In other words, it is taking a sample group that matches the survey category and is easy to survey.

1.3.3. Sampling Methods

In sample surveys, different methods of sampling are employed to obtain the required information. These are probability sampling methods and non-probability sampling method.

1.3.3.1 Probability Sampling Methods

In probability sampling methods, each method attempts to select samples such that each has a definable probability of being chosen. This approach includes, random sampling (simple, systematic and stratified random sampling) and, cluster sampling.

The procedures used to select a sample under this probability sampling methods, require that the sampler has some prior knowledge of the target population. This assists in the determination of the size of the sample needed to achieve a reasonable estimate (with accepted precision and accuracy) of the characteristics of the population. The general goal of all sampling methods is to obtain a sample that is representative of the target population. In other words, apart from random error, the information derived from the sample is expected to be the same had a complete census of the target population been carried out.

A random sample is one where every person (or unit) in the population from which the sample is taken has equal chance of being included in it. Ideally, the selections that make up the sample are made independently; that is, the choice to select one unit will not affect the chance of another unit being selected.

1.3.3.1.1 Simple Random Sampling

This is the simplest way of selecting sampling units where each unit has an equal probability of being chosen.

1.3.3.1.2 Systematic Random Sampling

This involves taking decision on what fraction of the target population is to be sampled, and then compiling an ordered list of the target population. The ordering may be based for instance on the date a student was admitted into school, the first letter of students' surname, age grade or bracket or other factors. Then, starting at the beginning of the list, the initial sample unit is randomly selected from within the first n units, and thereafter every n th individual is sampled. Typically, the integer or interval n is estimated by dividing the size of the target population by the desired sample size. This method of sampling is easy to implement in practice, and the sampling frame can be compiled as the study progresses. However it is especially vulnerable to periodicities in the list. If periodicity is present and the period is a multiple of n , then bias will result. It is important that the first name chosen is not simply the first in the list, but is chosen to be (say) the 5th, where 5 is a random integer in the range 1...10-1.

1.3.3.1.3 Stratified random Sampling

In stratified random sampling the target population is divided into distinct non-overlapping subgroups (strata) according to some predetermined characteristics (e.g., age, income) and then a random sample is selected within each subgroup. The sampler can use this method to ensure that each subgroup of interest is represented in the sample. This method generally produces more precise estimates of the characteristics of the target

population, except only where a very small number of units are selected within individual subgroup.

The two main reasons for using a stratified sampling design are firstly to ensure that particular groups within a population are adequately represented in the sample, and secondly, to improve efficiency by gaining greater control on the composition of the sample. In the second case, major gains in efficiency (either lower sample sizes or higher precision) can be achieved by varying the sampling fraction from stratum to stratum. The sample size is usually proportional to the relative size of the strata. However, if variances differ significantly across strata, sample sizes should be made proportional to the stratum standard deviation. Disproportionate stratification can provide better precision than proportionate stratification. Typically, strata should be chosen to have means which differ substantially from one another and which, minimise variance within strata and maximize variance between strata.

1.3.3.1.4 Cluster Sampling

This may be used if the study units form natural groups or if an adequate list of the entire population is difficult to compile. In a national survey, for example, clusters may comprise individuals in a localised geographic area. The clusters or regions are selected, preferably at random, and the persons are enumerated in each selected region and random samples are drawn from these units of the population.

Cluster sampling is an example of 'two-stage sampling' or 'multistage sampling'. In the first stage, a sample of areas is chosen; in the second stage a sample of respondent within those areas is selected. This can reduce travel and other administrative costs. It also means that one does not need a sampling frame for the entire population, but only for the selected clusters. Cluster sampling generally increases the variability of sample estimates above that of simple random sampling, depending on how the clusters differ between themselves, as compared with the within-cluster variation.

It is important to observe however that in any inquiry, a variety of sampling methods can be employed, individually or in combination.

1.3.3.2. Non-Probability Sampling Methods

This method of sampling is non-structured and, the probability of being included in the sample cannot be determined. Inclusion therefore is purely by chance. Examples of this sampling method include convenience sampling (sometimes called quota sampling) such as interview of passersby on the street by journalists and, volunteers for such things as testing of new products. These types of samples are prone to bias and cannot be assumed to be representative of the target population. For example interviewers might be tempted to interview those who look most helpful. Furthermore, people who volunteer are frequently different in many respects from those who do not. There is no way to assess the validity of the samples obtained using non-probability sampling strategies. In other words, tests of hypothesis and statistical inference concerning the sampled units and the target population cannot be applied. Under this method of sampling, there is no measure of the associated sampling error.

The distribution of values in any sample obtained by the probability sampling methods will differ from the distribution of values in sample chosen by chance or non-probability sampling method alone. The larger the sample in the former method, the more likely it is that the sample reflects the characteristic of interest in the target population. However, there are sources of error not related to sampling that may bias comparisons between the sampled units and the target population. First, coverage error (selection bias) may arise when the sampling frame does not fully cover the target population, in other words, the use of inadequate frame. Second, non-response bias may occur when sampled individuals cannot be reached or will not provide the information requested. Bias is present if respondents differ systematically from the individuals who do not respond. There is also the error associated with poorly designed questionnaire. Finally, the measuring device may not be able to accurately determine the characteristics being measured thereby resulting in measurement and recording errors. Never the less, once the sample has been taken the sampling error associated with the measured results can be computed.

From our preceding expositions of the key words in the topic, we can therefore, with a large measure of confidence, state that social survey method refers to a regular, systematic and scientific technique or procedure for the observation, collection and analysis of information or data on human society, its organisation and interaction for the purpose of understanding phenomena, predicting occurrences and intervening to promote or arrest the further happenings of the phenomenon.

1.4 DESIGN

Design is used both as a noun and a verb. As a verb, "to design" refers to the process of originating and developing a plan for a product, structure, system, or component. As a noun, "a design" is used for either the final (solution) plan (e.g. proposal, drawing, model, description) or the result of implementing that plan (e.g. object produced, result of the process). In philosophy, the abstract noun "design" refers to a pattern with a purpose. Design therefore refers to something that is not frivolous. It involves deliberate thought and planning with an anticipated aim which guides action. It has as its driving force the goal and objectives. In other words the designing of a survey instrument is dictated by the goal and objectives of the activity which has necessitated the need for the instrument.

1.5 INSTRUMENT

An instrument is a device that requires skill for proper use. It could also be viewed as the means whereby some act is accomplished. Instrument takes on different characteristics depending on the area of its application or use. It could mean method or tool for gathering or capturing and organising data from sample plus all the information and documentation that supports its use such as survey forms, questionnaires and observation forms. It could also mean a physical or electronic document that has a monetary value or records a financial transaction that determines trading rights between two parties such as a financial instrument. An instrument could also be a document in writing to create, alter, modify or terminate a right such as a legal instrument or, any device that produces a musical sound as in musical instrument. It could also mean a tool or device for measuring the value of an observable parameter or shows the extent, amount, quantity or degree of something such as a measuring instrument in the physical sciences.

2.0 DESIGN OF SURVEY INSTRUMENT

Generally in research, four types of research designs are employed. These are Experimental Research Design, Quasi-Experimental Research Design, Survey Research Design and, Case Study Research Design.

In experimental research design, two groups are used. These are the Experimental or Target Group and the Control or Check Group. For example if a particular text book is the major reference for a particular subject, a researcher may be interested in investigating the influence of the use of the book on students' performance level in that subject. In order to do this, the researcher uses the book to teach the experimental group, while the control group is taught with every other book but the reference text book. Before the commencement of the experiment, a pretest in the subject is conducted on the two groups. The two groups are then kept very far apart from each other, say, in different towns to prevent contact. Again after the experiment a post test is conducted on the two groups. If the experimental group performs better in the post test, then it is a proof that the reference book has an impact on performance in the subject. However if there is no difference in the performance of the groups when post tested, then we can conclude that the book does not really influence performance. This is called Significance.

Quasi-experimental research design is similar to the experimental but here, the barrier between the two groups is relaxed, that is, they are not as kept far apart. Survey research design has been dealt with in-depth elsewhere in this paper and, there is no need for a repetition. Suffice to add however, that surveys may be cross sectional where the respondent is interviewed only once and, longitudinal where the respondent is interviewed more than once. Case study research design is similar to survey and, is used for in-depth research study where results can be generalised.

As earlier observed, surveys take many forms, but often use questionnaires and interviews. Our focus in this section therefore is on the design of the various forms of social survey instruments including questionnaires, checklist and, interviews but more on questionnaires. The various forms of instruments are for purposes of generating data from an entity of interest. It is therefore necessary to consider the types of data that can be generated in a research enterprise.

2.1 Types of data

There are two types of data or random variables; categorical and numerical. Categorical random variables yield specific responses such as 'yes' or 'no'. However it is possible for categorical variables to yield more than two possible responses. For example in response to the question, 'On which day of the week are you most likely to wash clothes?' a respondent may mention more than one day in the week. Compared to the question, 'do you wash clothes on Sunday?' which will yield a categorical 'yes' or 'no' answer.

Numerical random variables yield numerical responses, such as the question, 'what is your height in centimeters?' Numerical variables are of two types – discrete and continuous. Discrete numerical random variables produce numerical responses from a counting process such as in the question 'how many times do you visit dentist in a typical month?' Continuous numerical random variables produce responses from a measuring process.

Height is an example of a continuous numerical random variable because the response takes on a value from an interval.

2.2 Sampling and Data Collection

Good data collection involves,

- Following the defined sampling process
- Keeping the data in time order
- Noting comments and other contextual events
- Recording non-responses

2.3 Review of Sampling Process

After sampling, it is advisable that a review of the exact process followed in the sampling, rather than that intended is held, in order to study any effects that any divergences might have on subsequent analysis. A particular problem is that of non-responses. Non response in survey sampling refers to a situation in which many of the individuals earlier identified as part of the sampling frame might have changed their minds from participating or may have indeed relocated from the target population. This leads to selection bias in conclusions.

The problem of non-participation is solved by follow-up attempt to contact the unresponsive and to characterise their similarities and differences with the rest of the frame. However for the samples that have relocated, this poses some problems which, may be surmounted by substituting the quantum of non-responses in the sampling frame with new samples selected through the same sampling method.

2.4 Design of Surveys Instruments and Survey Management

In social surveys generally, where the sampling units are usually human beings, the main methods of collecting information (survey instrument) are,

- Face-to-face interviewing
- Postal surveys
- Telephone surveys
- Direct observation
- The Internet

To these may be added 'enumeration' since social surveys may also involve the counting of objects such as vehicles and roads or the numbering of social and economic facilities such as houses, hospitals, markets, schools and the like.

Face-to-face Interview – This is a method of survey in which the interviewer is in direct physical contact with the respondent. Here the administrator holds a questionnaire from where questions are fielded to the respondent and the administrator gets his answers on the spot. This method of interview has the advantage of allowing the interviewee an opportunity to observe the respondent's demeanour or reaction to questions asked. It also provides opportunity for clarification of questions and answers. Answers obtained from this method have a high level of accuracy and non-response is virtually eliminated. However it has the disadvantage of either the respondent playing to the gallery in the answers or the interviewer influencing the response towards a desired direction. It is

therefore easy to introduce bias into the whole process which will bring errors and affect the result.

Postal Surveys – This involves mailing of questionnaires to respondents who are then required to mail back their responses. This method has the advantage of a wider reach as the sample population can be covered quickly. It however has the disadvantage of very large non-response. A way out of this is for the interviewer to prepay the return mail and follow up with telephone contact. This makes the this method more expensive than the other methods

Telephone Surveys – This is similar to face-to-face interviewing except that the both interviewer and respondent are not in direct physical contact. The questions-and-answers session is conducted over the telephone. This method has advantages similar to that in face-to-face interviewing. It also has the added advantage that the respondent will be more relaxed and less inhibited to answer questions. The method also has disadvantages similar to those in face-to-face interviewing. It however has the added disadvantage of the respondent being distracted by other things which the interviewer cannot see, breaking off the interview midstream, high cost of long telephone conversation or the telephone itself losing clarity thereby distorting the questions and answers with the inevitable undesired result.

Direct Observation – This is usually the method of survey employed in the physical sciences and medicine. It involves the physical observation of the subject and the taking of readings or measurements which are sometimes compared with some set parameters and from which results are extrapolated. However this method is also employed in the social sciences such as in the measurement of weather and climatic parameters, in sociology and social psychology.

Internet – This is the latest in methods of survey. The internet is the technological super highway which provides a treasure trove of information on virtually all subjects under the sun.

Enumeration – This is a survey method in which simple arithmetical counting and tallying of results are used to generate information which are then analysed, result obtained and inferences deduced. This method is usually employed in the survey of objects and when properly planned, errors are reduced to the barest minimum as they are least influenced by the vagaries of the human emotions.

Before choosing and embarking on the design of a survey instrument, it is important for the researcher to ask himself or herself the following main questions,

- What is the purpose of the survey?
- What kinds of questions would the survey be developed to answer?
- What does the researcher intend to do with the results obtained from the analysis of the data generated by the survey?

Whatever the method of survey chosen, a clear resolution of each of these questions will enable the researcher to determine what method of survey to use and this will in turn inform the structure or design of the survey instrument. A wrong or inadequate diagnosis of the purpose of the survey will obviously lead to the framing of inappropriate questions and, the receipt of faulty answers. This will in turn affect the result obtained from the analysis of the answers and, thus the deductions from such results. While this may not lead to serious negative consequences in social science enquiries, the result may be catastrophic in the area of epidemiology and medicine in which human lives are involved.

For the purpose of our topic, we have chosen to undertake an in-depth discussion of the questionnaire survey instrument. The choice is deliberate because experience has shown that this is the most commonly employed method of generating information in sample surveys in social research. Questionnaires are a series of structured questions designed to elicit the desired response from the respondents or the surveyed.

In designing the questionnaire, it is desirable and advisable to observe the following steps,

Step 1: Planning Questionnaire Research

- Consider the advantages and disadvantages of using questionnaires.
- Prepare written objectives for the research.
- Have your objectives reviewed by others.
- Review the literature related to the objectives.
- Determine the feasibility of administering your questionnaire to the population of interest.
- Prepare a time-line.

Step 2: Conducting Item Try-Outs and an Item Analysis

- Have your items reviewed by others.
- Conduct "think-aloud" with several people.
- Carefully select individuals for think-aloud.
- Consider asking about ten individuals to write detailed responses on a draft of your questionnaire.
- Ask some respondents to respond to the questionnaire for an item analysis. In the first stage of an item analysis, tally the number of respondents who selected each choice. In the second stage, compare the responses of high and low groups on individual items.

Step 3: Preparing a Questionnaire for Administration

- Write a descriptive title for the questionnaire.
- Write an introduction to the questionnaire.
- Group the items by content, and provide a subtitle for each group. Within each group of items, place items with the same format together.
- At the end of the questionnaire, indicate what respondents should do next.
- Prepare an informed consent form, if needed.
- If the questionnaire will be mailed to respondents,
 - Avoid having your correspondence look like junk mail.

- Consider including a token reward.
- Write a follow-up letter.
- If the questionnaire will be administered in person, consider preparing written instructions for the administrators.

Step 4: Selecting a Sample of Respondents

- Identify the accessible population.
- Avoid using samples of convenience.
 - Simple random sampling is a desirable method of sampling.
 - Systematic sampling is an acceptable method of sampling.
 - Stratification may reduce sampling errors.
 - Consider using random cluster sampling when every member of a population belongs to a group.
 - Consider using multistage sampling to select respondents from large populations.
- Consider the importance of getting precise results when determining sample size.
- Remember that using a large sample does not compensate for a bias in sampling.
- Consider follow up sampling of non respondents to get information on the nature of a bias.
- Remember that using a large sample does not compensate for a bias in sampling.
- Consider sampling non-respondents to get information on the nature of a bias. The bias in the mean is the difference of the population means for respondents and non respondents multiplied by the population non-response rate.

Step 5: Preparing Statistical Tables and Figures

- Prepare a table of frequencies.
- Consider calculating percentages and arranging them in a table with the frequencies.
- For nominal data, consider constructing a bar graph.
- Consider preparing a histogram to display a distribution of scores.
- Consider preparing polygons if distributions of scores are to be compared.

Step 6: Describing Averages and Variability

- Use the median as the average for ordinal data.
- Consider using the mean as the average for equal interval data.
- Use the median as the average for highly skewed, equal interval data.
- Use the range very sparingly as the measure of variability.
- If the median has been selected as the average, use the inter-quartile range as the measure of variability.
- If the mean has been selected as the average, use the standard deviation as the measure of variability.
- Keep in mind that the standard deviation has a special relationship to the normal curve that helps in its interpretation.
- For moderately asymmetrical distributions the mode, median and mean satisfy the formula: $\text{mode} = 3 * \text{median} - 2 * \text{mean}$.

Step 7: Describing Relationships

- For the relationship between two nominal variables, prepare a contingency table.
- When groups have unequal numbers of respondents, include percentages in contingency tables.
- For the relationship between two equal interval variables, compute a correlation coefficient.
- Interpret a Pearson r using the coefficient of determination.
- For the relationship between a nominal variable and an equal interval variable, examine differences among averages.

Step 8: Estimating Margins of Error

It is extremely difficult, and often impossible, to evaluate the effects of a bias in sampling. However the following should be noted;

- When evaluating a percentage, consider the standard error of a percentage.
- When evaluating a mean, consider the standard error of the mean.
- When evaluating a median, consider the standard error of the median.
- Consider building confidence intervals, especially when comparing two or more groups

Step 9: Writing Reports of Questionnaire Research

- Use the median as the average for ordinal data.
- Consider using the mean as the average for equal interval data.
- Use the median as the average for highly skewed, equal interval data.
- Use the range very sparingly as the measure of variability.
- If the median has been selected as the average, use the inter-quartile range as the measure of variability.
- If the mean has been selected as the average, use the standard deviation as the measure of variability.
- Keep in mind that the standard deviation has a special relationship to the normal curve that helps in its interpretation.
- For moderately asymmetrical distributions the mode, median and mean satisfy the formula: $\text{mode} = 3 * \text{median} - 2 * \text{mean}$.

2.5 Questionnaires Administration

When using the questionnaire whether in the face-to-face interview or mailed (postal) method, it is important to make the respondent feel at home when answering the questions if the right answers are to be obtained. In designing the questionnaire therefore the following tips come in handy,

- As much as possible, assure people that the surveys are anonymous
- Explain to the respondent the reasons for asking the questions,
- Make response categories as broad as possible.
- Word the question in a non-judgmental style that avoids the appearance of censure, or, if possible, make the behavior in question appear to be socially acceptable.
- Present the request as factual matter as possible.
- Guarantee confidentiality or anonymity

- Make sure the respondent knows the information will not be used in any threatening way.
- Explain how the information will be handled
- Avoid cross classification that will allow for pinpointing responses.

3.0 STATISTICAL METHODS

The following are some of the statistical methods employed in the analysis of data obtained from surveys;

1. The **mode** is the most frequently occurring score in a data set. Example: In the following data set 3, 12, 14, 16, 17, 18, 19, 22, 22, 22, the mode is 22.
2. The **median** is the score that divides a data set in two (the score in the middle). Half the scores are higher than the median, and half are lower than the median. Example: In the data set, 3, 12, 14, 16, 17, 18, 19, 22, 22, 22, 22, the median is 18. When there is an even number of scores in a data set, the median is halfway between the two middle numbers.
3. The **mean** is the arithmetic average obtained by adding the numerical values of all the scores in a data set and dividing that sum by the number of scores in the data set. Example: In the data set, 3, 12, 14, 16, 17, 18, 19, 22, 22, 22, 22, the mean is equal to $(3 + 12 + 14 + 16 + 17 + 18 + 19 + 22 + 22 + 22 + 22)/11 = 187/11 = 17$. It therefore means that an extreme score can change the mean radically, possibly making it less representative of the data.
4. The **range** is a measure of variability computed by subtracting the lowest score from the highest score in a data set. The range is affected by extreme scores. Example: In the data set 2, 3, 4, 5, 5, 5, 6, 7, 8, 100, the range is $100 - 2 = 98$. If the extreme score (100) is dropped, the range is $8 - 2 = 6$. Extreme scores can therefore radically affect the range of a data set.
5. The **standard deviation** is also a measure of variability. It reflects the average distance between each score and the mean of a data set. The standard deviation shows how different the scores in a data set are from the mean. Example: Following are two data sets; 1st Set: 1, 2, 3, 4, 5, 6, 7, 8, 9 and, the 2nd Set: 4, 4, 4, 4, 5, 6, 6, 6, 6. the mean of both data sets is 5. However, the scores in the 1st set are a greater distance from the mean. In other words, they are more different from the mean than the scores in the 2nd set. Therefore, the standard deviation (SD) in the 1st set is larger than the SD for the 2nd set.

6. A **correlation** is an indication of the relationship between two variables (x and y). As x decreases or increases so does y correspondingly decreases or increases. In other words x could negatively or positively correlate with y. However correlations do not indicate causation.

7. The **correlation coefficient** (r), a number between -1.00 and +1.00, is a mathematical representation of the strength and direction of a correlation. The higher the absolute value of r is, the stronger the relationship is. A perfect correlation, whether positive or negative (where r equals + or -1.00), describes a perfect relationship; knowing the value of x allows the certain prediction of y. A positive correlation (where r varies from 0 to +1.00) describes two variables that change in the same direction: as x increases, so does y (and vice versa). A negative correlation (where r varies from -1.00 to 0) describes an inverse relationship: as x increases, y decreases (and vice versa).

8. **Inferential statistics** are used to judge the meaning of data. Inferential statistics assess how likely it is that group differences or correlations would exist in the population rather than occurring only due to variables associated with the chosen sample.

9. **Statistically significant**, refers to a situation in statistics which indicates that the group differences or correlation is larger than would occur by chance. Example: If the difference between the means of two groups of data is statistically significant, a researcher would conclude that the difference most likely exists in the population of interest. If the difference is not statistically significant, a researcher would conclude that the difference occurred by chance, possibly because of an unrepresentative sample or the presence of confounding variables.

REFERENCES

- Bartlett J.E., Kotrlik J.W. and Higgins C. (2001), Organisational Research: Determining Appropriate Sample Size in Survey Research. *Journal of Information Technology, Learning, and Performance*, 19(1) 43-50
- Bradburn, N. M., and Sudman, S. (1988). *Polls and Surveys: Understanding What They Tell Us*. San Francisco: Jossey-Bass Publishers.
- Brown, K.W., Cozby, P.C., Kee, D.W., & Worden, P.E. (1999). *Research Methods in Human Development*, 2d ed. Mountain View, California : Mayfield. ISBN 1-55934-875-5
- Crawford S, Stucki L (1990), "Peer Review and the Changing Research Record", "Journal of American Social Information Science", vol. 41, pp 223-228
- Encyclopedia of Public Health, The Gale Group, Inc.
- Festinger, L., and Katz, D., (eds). (1966). *Research Methods in the Behavioral Sciences*. New York: Holt, Rinehart and Winston.
- Glen, William (ed.), *The Mass-Extinction Debates: How Science Works in a Crisis*, Stanford University Press, Stanford, CA, 1994
- Kelsey, J. L.; Thompson, W. D.; and Evans, A. S. (1986). *Methods in Observational Epidemiology*. New York: Oxford University Press.
- Morris, J. N. (1975). *Uses of Epidemiology*, 3rd edition. London: Churchill Livingstone.
- Moser, C. A., and Kalton, G. (1989). *Survey Methods in Social Investigation*, 2nd edition. Aldershot, UK: Gower Publishing Company.
- Pagano, M., and Gauvreau, K. (2000). *Principles of Biostatistics*, 2nd edition. Pacific Grove, CA: Duxbury.
- The American Heritage Dictionary of the English Language, Fourth Edition 2007, 2000, Houghton Mifflin Company.
- The Concise Oxford Dictionary of Archaeology. 2002, 2003 Oxford University Press.
- The Veterinary Dictionary, 2007, Elsevier Publishers.
- US General Accounting Office book, "Developing and Using Questionnaires" (Oct 1983) chapter 9,