TECHNIQUES FOR GATHERING SOCIO-CULTURAL DATA ON WATER SUPPLY AND SANITATION PROGRAMMES

A TECHNICAL GUIDE

COMPILED BY
BABY S. MOGANE

EDITED BY
B.J. DIXON &
M.A. TROLLIP

Water
Technology

CSIR

DIVISION OF WATER TECHNOLOGY CSIR

TECHNIQUES FOR GATHERING SOCIO-CULTURAL DATA ON WATER SUPPLY AND SANITATION PROGRAMMES

by

S B Mogane

PRETORIA
MAY 1990

ABSTRACT

In recent years, there has been a shift in emphasis on rural development due to major setbacks caused by planners' reluctance to recognize social expertise in development projects in the past. Socio-cultural data has, therefore, become an essential feature of rural development projects. In this report, relevant qualitative techniques of gathering and analysing social information for water supply and sanitation projects in developing areas are discussed.

CONTENTS

TECHNIQUES FOR GATHERING SOCIO-CULTURAL DATA ON WATER SUPPLY AND SANITATION PROGRAMMES

1. INTRODUCTION

In their effort to alleviate poverty, major International Aid Agencies like the World Bank have for a long time invested a lot of capital for development projects throughout the Third World countries. Despite all these efforts, limited success was achieved. Water supply and sanitation programmes were no exceptions to the latter.

Subsequent evaluation studies revealed that development experts of the sixties and early seventies were less convinced than at present about the value of social expertise throughout the project cycle. According to Kottak (Cernea, 1985: 325), the development experts often perceived project participants as no more than a collection of people - potential beneficiaries rather than as structured groups of active individuals with their own perceptions of needs and desires to help plan and implement changes that will affect their own lives and those of ensuing generations. In addition, Uphoff (Cernea, 1985: 359) states that people have been identified as 'target groups' rather than 'intended beneficiaries'. All these mean that people have been left out of the design, therefore, dehumanized.

The widening gap between project planners and the people has led to the recognition of social expertise in development programmes including water supply and sanitation. With the inception of International Water Supply and Sanitation Decade (1981 - 1990), engineers have increasingly found themselves working in multi-disciplinary teams which may include sociologists, anthropologists, economists, health educators and others. In some cases, however, an engineer himself may be asked to gather social science data on the community to be provided with water and sanitation improvements. Therefore, it is useful for the engineer, development experts or even some social scientists who are newcomers in this field to have guide on the various types of data gathering methods used in social science (Simpson-Hébert, 1983: 1).

2. RATIONALE FOR SOCIAL INFORMATION

In contrast to urban communities, the provision of water supply and sanitation in developing communities is a complex issue and therefore, requires the involvement of intended beneficiaries for the following reasons:

- Developing communities often have access to water which is readily available from natural sources such as rivers and springs. To these communities, therefore, water is a free commodity whereas urban communities are accustomed to the fact that the convenience of their water supply should be paid for. Besides being accustomed to the payment, urban dwellers usually have an easy access to cash whereas their rural counterparts have little capacity to pay.
- Although water is a basic and felt need, its use has important health implications that may go unrecognized (Elmendorf and Buckles, 1980: 1). For many people, if water is clear, it is safe to drink. Therefore, there is often a need for an explanation of the germ theory to developing communities, so that they can appreciate the

need to change traditional water and sanitation practices if they are deemed inappropriate. As far as possible, designs for improved schemes should fit in with the traditional patterns of water use and excreta disposal. Moreover, the designs should meet local preferences and values (Simpson-Hébert, 1983: 3).

- Implementation of low-cost water supply and sanitation projects depends heavily on the use of local materials and expertise, therefore, co-operation of all members of the community is essential. Hence, planners need to have some knowledge of social and political structures in the community.

For these reasons, planners need to be adequately informed about the

cultural beliefs influencing traditional behaviour. Kottak (Cernea, 1985: 325) sums up this aspect by stating that social engineering is as important as technical or financial or economic considerations in project design. Therefore, to an extent possible, each project must have a socially informed and culturally appropriate design and implementation strategy.

This report does not claim to provide answers to all problems related to social science data gathering techniques. All it does is to provide basic guidelines based on previous field experience as well as on literature in water supply and sanitation.

3. KIND OF DATA NEEDED

The range of data needed for water supply and sanitation programmes may include the following:

3.1 Bio-demographical information

- Population size, growth rate and mobility
- An understanding of village organizational structures and identification of disadvantaged groups
- Establishing opinion leaders and influential persons within the community
- Types of dwellings, their physical condition and layout
- Family composition and size (e.g. nuclear or extended families)
- Age and educational qualifications of respondents

3.2 Water sources, usage and attitudes

- Available sources and purpose of use for each household (e.g. drinking, cooking, laundry, bathing, animals, home, garden)
- Quantity and quality
- Distance and time to fetch water
- Seasonal variations in water source
- Drawers of water (e.g. women, children)
- Preferred water sources for related activities (e.g. laundry, animal drinking, bathing)
- Household storage

- Perceptions of community needs
- Local beliefs and attitudes related to water sources

3.3 Health aspects

- Morbidity and mortality rate (especially among children under the age of five)
- Existing water related/borne/washed diseases
- Perceptions of diseases and cure
- Major health problems in the community and seasonal variations

3.4 Technological affordability and willingness to contribute

- Heads of households and major contributors to upkeep of families
- Means of subsistence or major occupations
- Preferred spending patterns and ability to contribute money
- Borrowing and savings customs
- Seasonal employment
- Indirect measurement of household incomes (e.g. expenditure patterns and monthly contributions by employed members of the families)
- Payment of water

3.5 Technological alternatives

- Local skills, capabilities and traditional alternatives
- Preferred technological improvements
- Availability of skilled and unskilled labour
- Local availability of materials for construction

3.6 Excreta disposal and relationship between method and health

- Existing defecation practices (noting important differences between religions, men, women, in-laws and children)
- Personal hygiene habits, cleansing materials and practices
- Important taboos, beliefs, related locations, sharing etc
- Latrine emptying, sludge reuse, rubbish and waste water disposal
- General household cleanliness

3.7 Community involvement and participation

- Major local organizations and type of membership
- Community and family level leadership in decision-making
- Major local political or social factions which might affect participation
- Extent of interest and participation in water, sanitation and other development activities

- Important characteristics that would determine acceptability and influence of outsiders working in the area
- Priority given to improvement of water supply and sanitation in relation to other priority needs in the community

3.8 Communication media

- Mass media access in the area
- Coverage by volunteers and field workers
- Effectiveness of different media for different tasks
- Audio-visual perceptions, literacy rates, language and dialect
- Ongoing formal or non-formal health education activities

3.9 Livestock

- Availability and number of livestock (e.g. cattle, sheep, goats)
- Sources of water for drinking
- Perceptions of needs as far as live-stock is concerned

4. ELIGIBILITY FOR INTERVIEWS

Effort should be made to interview a representative sample of all members of the community irrespective of their political affiliation or social status. The wider the variety of people interviewed, the greater the range of opinions will be established. There is often a tendency by urban based professionals to be involved in what Chambers calls rural development tourism. According to Chambers, rural development tourism introduces biases that work against perceiving rural poverty, reinforce underestimates of its prevalence, and prevents understanding of its nature (Cernea, 1985: 400). Often, the poorer people are out of sight of the road and live in inaccessible hamlets. Rural development tourism tends to meet the less poor and more powerful and men rather than women.

Despite the fact that in most cultures the task of fetching water is primarily women's, they are usually deprived of the opportunity to voice their opinions regarding the choice and siting of water schemes. Therefore, it is imperative that disadvantaged groups in the community including women receive attention in order to avoid domination of the elite in water organization. This view is taken further by Mills (1987: 165) when she states that dialogue and negotiation should be achieved with as wide a cross-section of the community as possible. She states that this is essential in order to avoid the nett effect. The nett effect refers to those who are powerful and likely to absorb and intercept development input before it reached all sections of the community. In her action research in Boschfontein, Kangwane, Mills had from the outset taken cognisance of the power structure and ensured that neither the traditional leaders, the professionals or business elite, or the youth leaders blocked the benefits of development for any other interest group. She therefore negotiated the viability of the project with all including the youth leaders who perceived themselves as political liberators for the community.

5. TECHNIQUES FOR GATHERING SOCIO-CULTURAL DATA

For a long time, methods of gathering social data have been overly time-consuming, elaborate and often the results were outdated. This is because social scientists allowed themselves endless months, if not years to study communities. Consequently, the results could not be utilized for implementation purposes, and were therefore dismissed. Often development projects, particularly those studying water supply and sanitation, lack the luxury of time for data collection and report writing. Although there is a need for cost-effective methods of gathering social data, planners should guard against what Chambers calls 'quick and dirty,' meaning a cheap and fast analysis (Cernea, 1985: 400-403). One should also guard against the case where the researcher spends many months on theoretical aspects and not necessarily answering the relevant question.

Fortunately in the past few years, sociological and anthropological researchers involved in development work have significantly refined their approaches and techniques. Currently, more simplified, time-efficient and cost-effective methods have been developed. There are various methods of gathering social information. However, just a proven few relevant for water supply and sanitation projects will be covered in this report. Most information contained in this section was elicited from the work of Simpson-Hébert (1983), Cernea (1985), Chandler (1986) and Smit (1987). We will now turn to the discussion of different techniques of gathering data.

5.1 Literature survey

There is often a wealth of information for those who take the trouble to look for it, and yet it is ignored by researchers/planners. Ignorance of such information often increase project cost because planners tend to start from scratch. Planners should adopt enquiring minds if projects are to be cost-effective and time-efficient. Unfortunately it is often difficult, if not impossible to trace some of the unpublished ad hoc research by university students, tutors, development experts and others particularly in self-governing and independent states of South Africa. However, efforts should be made to trace whatever information there is available. Time spent in searching for information even when it is not known whether it exists, is often well repaid and may avert grave errors as well as save unnecessary demands for new data collection. The USAID impact evaluation missions now deliberately devote time to studying sociological, economic and geographical literature before going to the field (Cernea, 1985: 405).

In South Africa, a wealth of valuable information is often available from university libraries, Development Bank of Southern Africa, Human Sciences Research Council, Department of Statistics and other relevant departments, development organizations, Government Printers and others. In most cases, these organizations can provide information such as maps, aerial photographs, population census, satellite imagery about ecology, settlement patterns, infrastructure and academic reports and research papers which can be invaluable to planners. Therefore, these resources should be adequately exploited when planning for a project.

5.2 Aerial inspection and surveys

Although a costly option, aerial inspection and surveys can be invaluable, particularly from helicopters. The utility of aerial surveys for counting animals and certain types of natural resource surveys is well established. Besides giving a general spatial perspective on an area, thus, making it possible in a matter of minutes to get a broad overview of land use, crop zoning and the like, aerial surveys can also help in identifying where the poor live - on the outskirts of villages in scattered hamlets and more remote places.

However, aerial inspection and surveys are less well recognized in their value for purposes close to people. This technique should, however, not become substitutes for work on the ground and for learning from people. An antidote is to fly with someone who is locally knowledgeable and who can interpret what is seen (Cernea, 1985: 410). Due to its limitation as far as contact with people is concerned, therefore, this technique needs to be used in combination with other techniques such as open-ended questionnaires still to be discussed.

5.3 Key Informant Interviewing

In any community there are people who are more knowledgeable about aspects of community life. Some of the most useful informants are social workers, health workers, tribal authority, agricultural extension officers, priests and school teachers. In most cases, however, key informants are those who are better off, educated and more powerful. Although this technique can provide a valuable information, it needs to be used with awareness of its limitations. In their study, Young and Young (Hébert-Simpson, 1983: 8) found that key informants were more reliable in giving information about the following:

- Physical geography and buildings (e.g. "Is there a health unit here?"

  "What is your main water source for drinking?")
- Institutions and institutional roles (e.g. "Do you have a latrine builder here?")
- Dates of important community events (e.g. "When did you get a well in this village?")

According to these researchers, there is usually much lower degree of agreement among informants when it comes to more evaluative questions such as "What percentage of people here would like to have a latrine?"

Another pitfall is that the information provided by key informants is biased, therefore needs to be consciously off-set by seeking those informants who tend to be left out and uncontacted. These people are the women and the poorer people, who are often much better informed and articulate than the outsiders are conditioned to expect (Cernea, 1985: 408). In essence what is said here is that key informants should also include the disadvantaged groups in the community. Naturally, a rapport with all people whose co-operation will be sought in the project such as local engineers, teachers, community workers, the disadvantaged groups, health and agricultural extension officers needs to be established. In this way a good working relationship is equally promoted through the process of gathering information.

In view of the fact that key informant interviewing is often biased, thus, unreliable, it is recommended that it be used in combination with either participant observation or a survey, due to their usefulness in eliciting reliable information.

5.4 Participant observation

Participant observation is one of the most basic and widely used social science method of gathering data. As a participant observer, the researcher establishes residence in the community to be studied and remains there for weeks or months observing and recording activities of daily life. The researcher is an active observer and gathers information relevant to the data needed (Simpson-Hébert, 1983: 6).

The researcher virtually gets involved in all activities of the community like fetching water, building houses, laundry etc. In addition he asks questions and all the time, detailed notes are kept from what is heard, seen and felt about the subject under study. Visits to the water sources and counting the number of trips made to collect water, how water is used and handled, toilet habits, washing and personal hygiene can yield a great deal of factual data.

Chambers (Cernea, 1985: 407) states that rural people often have beliefs about their values and activities which do not correspond with reality. Therefore, participant observation often reveals that the custom has either lapsed or was never practised. Through participant observation, a lot of essential information on interpersonal relations, political structures and authority figures could be elicited. Though time-consuming and therefore not an answer for rapid rural appraisal projects, participant observation is one of the most important ways of learning by doing.

5.5 Open-ended questioning

Open-ended questioning is a method of acquiring information from respondents without a formal interview schedule. This technique is normally applied when the researcher lacks reliable information about the community as well as the subject under study. In addition to eliciting information, an interviewer equally familiarizes himself with the area and its people in order to come up with structured questions, if a survey is deemed necessary. Though informal, a checklist of questions which an interviewer uses as a flexible guide is needed. Chambers states that this technique is an art, needs a sensitive balance between open-endedness and direct enquiry in order to identify questions the outsider does not know how to ask and yet covers the major concerns (Cernea, 1985: 409). Of equal importance again in this method is the establishment of rapport and creation of a relaxed atmosphere for respondents.

Open-ended questions are broad in nature, often specifying only the topic covered. This technique is respondent-centred rather than interviewer- centred. Another way of administering open-ended questions is through what Simpson-Hébert (1983: 9) calls "heuristic elicitation". Heuristic elicitation simply means that the next question is based on the answer to the previous question. As a result, the questions are respondent-generated rather than interviewer-generated, thus are likely to be comprehensive. For example, instead of asking straight forward questions like "What type of a latrine would you prefer?" "What type of superstructure should it have?" "Where should it be located?", the interviewer should ask the respondent to describe the attributes of all the places where he habitually urinates and defecates and then how he feels about each one relative to the other attributes.

Other examples of open-ended questions are:

- Tell me about your water supply situation
- Describe the sanitary facilities available to children
- Tell me about the water supply technologies you know of and how you feel about each of them

In view of the broadness of answers to these questions, adequate space should be provided on an answer sheet to accommodate all information provided. Even better, a tape recorder could be used. However, prior permission to record the conversation should be obtained from the respondent. Open-ended questions should be administered to a minimum of thirty people, this should provide an adequate range of responses. In water supply and sanitation projects, it is important to interview men, women and children as the project must be designed to reach all categories (Simpson-Hébert, 1983: 10).

Lastly, the data should be analysed. In view of the fact that open-ended questions are broad, therefore, difficult to code in advance, the researcher needs to categorize responses and provide codes for analysis. If time is short, the project can proceed based on this information combined with either methods discussed above.

5.6 Group interviews

Group interviews are the most cost-effective way of gathering a variety of information in a short space of time with manageable clusters of five or six people. They are especially useful in acquiring information on natural resources, when a wider geographical area and subject matter can be covered than with one respondent. Besides gaining a larger body of knowledge, group interviews also provide a chance for mutual checking (Cernea, 1985: 409). The current trend among black communities in South Africa is the mushrooming of small interest groups which may include burial societies, church, youth, political and recreational groups. In most cases these groups are constituted by people from different educational levels and social status. All the researcher needs to do is to identify these groups and convene panels for discussions. As with other methods discussed above, rapport and a relaxed atmosphere are essential ingredients for success. For example, Chambers conducted group interviews with small clusters of women of two or three generations on patterns of infant feeding. It was established that there was a self-correcting mechanism within a group. If one person puts across an overfavourable picture of her own group's behaviour, a peer would give a more realistic observation (Cernea, 1985: 409).

In our (CSIR) recent water supply video evaluation study with the Human Sciences Research Council in Inanda, an informal settlement area near Durban, a few group interviews with various groups were successfully conducted (unpublished report, 1989). Our success was brought about by the more relaxed and friendly atmosphere established with all groups as well as rewards in terms of refreshments and/or nominal amount of money at the end of each session. All sessions were most revealing and enjoyable. Therefore, a lot can be learned from group interviews.

5.7 Surveys

Conducting a survey is the most commonly used method of gathering social science data. With surveys, observation in one way or another is essential. The commonplace instruments used in gathering data are questionnaires or interview schedules with structured questions to be administered to a sampled population. It is imperative that the questionnaire or an interview schedule be cautiously designed in order to elicit precisely what the researcher wishes to learn. In other words, simple language and courtesy in choosing words should be exercised when designing these instruments.

Surveys are useful in providing statistical evidence on details such as household incomes, expenditure patterns, morbidity and mortality rate, educational levels, and the like based on a sample. Although the most useful method of collecting data for systematically quantifying the occurrence of observable objects and for estimating the prevalence of particular attitudes, beliefs and values, surveys have several drawbacks which include

- time inefficiency
- high cost
- use of pre-structured questions which tend to limit exploration into cultural beliefs and values (Simpson-Hébert, 1983: 10)

However, when properly designed and based upon an earlier heuristic elicitation study, surveys can be invaluable. For example, Mills (1987: 163) negotiated the viability of a water and sanitation improvement project, first of all with the tribal authority and representatives of professional as well as business elites. These community leaders agreed that the community needed water and sanitation. It was further agreed that the needs be identified by means of collaborative dialogue. Therefore, a formal meeting with the community to discuss the proposal was convened. The community unanimously responded in favour of the project. During the ensuing dialogue, it was established that there was a definite shortage of water and that latrines were very unpleasant, therefore not regularly used. Although the researcher had enough data to proceed with the project, she found it necessary to conduct a limited survey to ensure that specific local details were ascertained in order to quantify the data, hence, conducted the survey.

Surveys may either be based on the total population (often called census) or based on samples drawn from a population. However, in small communities with households numbering less than 200, a census is possible, but in larger communities complete coverage is impractical. The advantage of a census is that it can provide the statistical confidence gained from having surveyed all households. However, if budget, manpower and time do not permit, complete survey is not necessary, therefore, sampling is preferable. It is on this point that we turn to sampling.

6. SAMPLING

A sample is anything less than a full survey of a population. It is usually thought of as a small part of the population, taken to give an idea of the quality of the whole (Hannagan, 1982: 43). There are various sampling methods to choose from in carrying out a sample survey.

6.1 Simple random sampling

Simple random sampling is a procedure by means of which each unit (the household usually the unit of sampling in water and sanitation projects) has an equal chance of being selected. This method assumes that one has been able to identify every unit in the universe to be sampled and that every unit is represented. A lottery method can be used as a basis of selecting each unit. In the lottery method the population is arranged sequentially and assigned a numerical identification. Corresponding numbers are marked on separate tabs and put into a revolving drum or closed container. The numbers are tossed so that they are thoroughly intermixed. Then one tab bearing a number is selected from the total number of tabs in the container without the selector seeing the tab pool. The number is recorded and the tab is then tossed back into the tab pool. The same procedure should be used until the required total has been reached. This is an important feature of the lottery method. It ensures that every individual has the same chance of being chosen as every other individual (Leedy, 1974: 94).

However, for large groups, it is not possible to number each unit of the universe and pick them out randomly. In this case, therefore, a table of random numbers (which can be found as an appendix in nearly all statistics textbooks) can be used. A table of random numbers consists digits 0 - 9 (usually in groups of 4 or 5 as indicated in Figure 1) in which each figure occurs approximately the same number of times and the occurrence of any figure or number is entirely independent of the occurrence or not of any other figure in the table.

Hannagan (1982: 49) gives this example: In random sampling for consumer market research, the Electoral Register is often used as a sampling frame (eg. maps, list of houses, names etc.). The electors for the polling district are listed according to name and address, and each elector is given a number. If a particular register has 5000 electors, each elector will have a number from 1 to 5000. Random numbers can be read off the table in groups, for instance, 2412, 8627, 0143. The first sample number would be elector 2412 on the electoral register, the second elector picked would be number 0143. Random number 8627 would be ignored because the sample population is only 5000. This process is continued until the number of people required for the sample is reached (say 500 for a 10 % sample).

Another way of drawing a random sample if no enumeration is possible is by using aerial photographs. Aerial photographs are particularly useful in rural areas with scattered households which do not form a grid-like pattern. The procedure is to number every house in the area on a photograph. Numbers are then selected from a table of random numbers and then the houses corresponding to these numbers circled on the aerial photograph (Simpson-Hébert, 1983: 11). Again this process can be continued until the required total of households is reached. The schematic process of simple random sampling is thus (Leedy, 1974: 102):

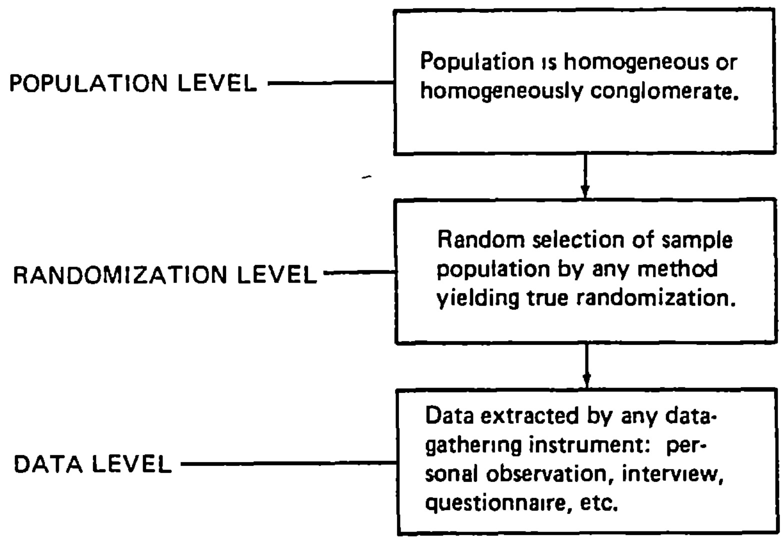


Figure 1: *Simple random sampling design*

6.2 Systematic sampling

Systematic sampling implies the random selection of certain items in a series according to a predetermined sequence. First of all, a decision must be made about the number of elements we need in the sample, calculate what proportion of the sample is of the universe, and then use this proportion as the sampling interval. If the sampling proportion is 10 %, then the sampling interval is every tenth unit. To select every tenth element randomly, all elements of the universe must be arranged in some consistent systematic order. For example, if we were to sample every tenth household in a town, we would simply start with one house and move through space in a systematic pattern taking every tenth house in the whole universe. The advantage of this method is that we do not have to identify every dwelling unit in advance. The basic disadvantage is that we must know in advance the size of the sample needed and proceed systematically through the whole universe before the sample is representative (Gordon, 1980: 238).

Leedy (1974: 106) gives a schematic example of systematic random sampling thus:

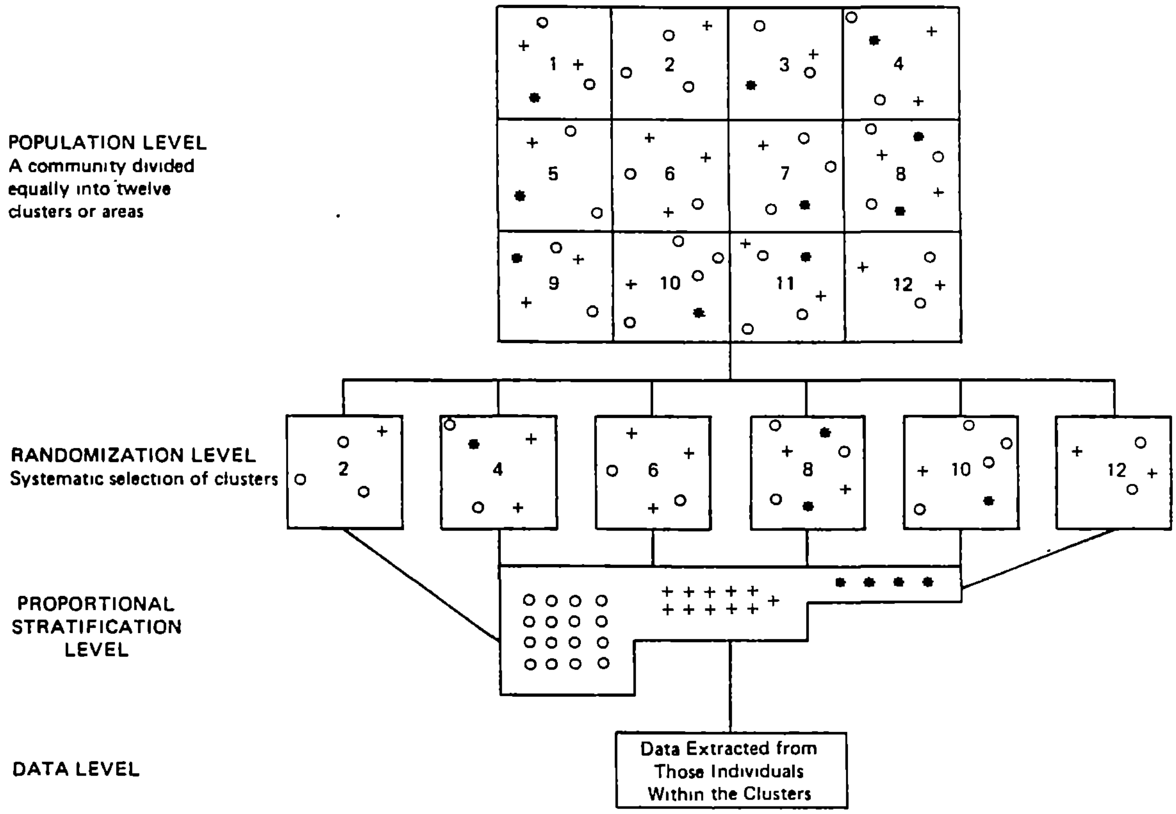


Figure 2: *Systematic random sampling*

6.3 Stratified random sampling

In stratified sampling all the people or items in the sampling frame are

divided into groups. The strata are identified by certain attributes known to be related to the opinions, beliefs or attitudes we are trying to sample. For example, if we want to sample attitudes of the community towards their water supply and sanitation we could stratify the universe into sub-samples by age, sex, educational qualifications, political affiliation and the like. However, the more we know about respondents in the universe and how characteristics correlate with the opinions we are trying to measure, the more effectively we can stratify the sample and gain precision without increasing the number of respondents in the sample (Gorden, 1980: 245).

For example: In a marketing survey the sales of cigarettes in a variety of outlets may be investigated by dividing the retail outlets into strata. In a particular town or village shops may be divided into large, medium and small outlets and a simple random sample taken based on shops from each category. A clear definition of the strata is important so that there is no overlap between shops. (Hannagan, 1982:51)

6.4 **Quota sampling** (Non-probability sampling)

In quota sampling a certain number of people with specific characteristics are interviewed. It is stated (Smit, 1987: 83) that it is difficult to say just what quota sampling is, what it does and what it is supposed to do. However, the following applies to most quota samples. The sizes of sub-classes (formed by control variable such as sex, age and geographical area) are estimated from census data or other available data. The quotas of the number of sample elements in each sub-class are then determined proportionally to the population sizes in the sub-classes. These sample quotas are then divided among the interviewers who attempt to find persons who satisfy the stated requirements. Obviously the method according to which sample elements are obtained and the extent to which fieldwork is supervised can have an important influence on the quality of the results.

Hannagan (1982: 52) gives this example: interviewers may be told to

interview, over a period of several days, fifty people divided into age and socio-economic groups, to ask them their opinions on a television advertisement. These groups may be divided into proportion to the numbers in the population. What the interviewers do then is to interview the people who fit the characteristics of the respondents.

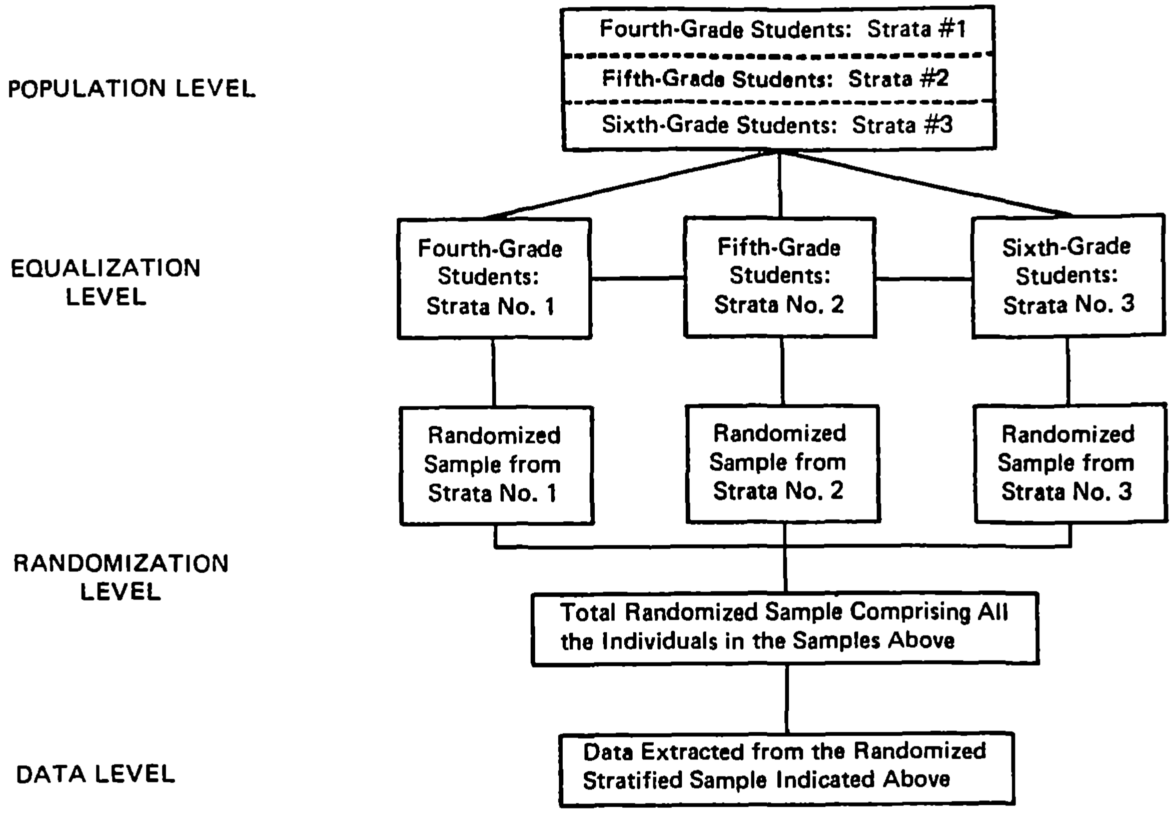


Figure 3: *Stratified random sampling*

The characteristics of random, systematic and stratified sampling are that every individual has a known probability of being included in the sample.

6.5 Cluster sampling

In cluster sampling (or area sampling) clusters are formed by breaking down the area to be surveyed into smaller areas, a few of these areas are then selected by random methods and units (such as households) are interviewed in these selected areas. The units are selected by random methods.

For example: In an impact evaluation study of water supply (spring protection) project in the Valley of a Thousand Hills, near Durban, a cluster sampling technique based on approximately sixty protected springs was used (Mogane, Internal report, 1987: 6). The rationale for using this technique was due to the vastness of the area with dispersed population and the fact that only sixty springs were protected in the whole area. Therefore, using a table of random numbers, eighteen springs were selected for the survey and a few randomly selected households around each spring were interviewed. Cluster sampling is convenient to use in field work and reduces costs but is less precise than simple random sampling.

Leedy (1974: 104) gives a schematic example of cluster sampling thus:

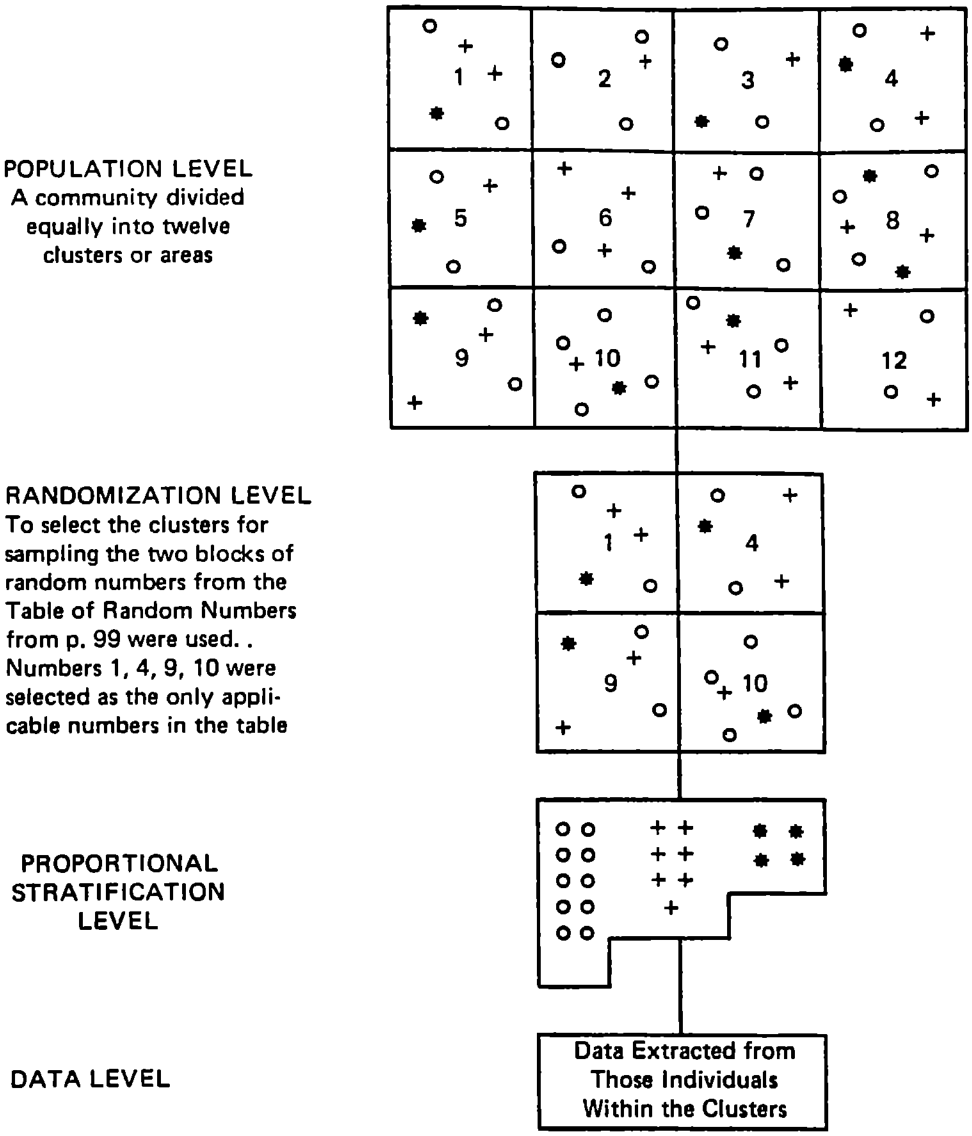


__Figure 4:__ Cluster or area sampling

There are many other methods of sampling available in survey and statistical textbooks but have not been covered in this report. For further reading it is recommended that the Human Sciences Research Council's report entitled 'Survey Methods of Practice' by P C Smit (1987) be consulted.

Lastly we shall briefly discuss the sample size. Sample size depends largely on the characteristics and qualities of the population. The more heterogeneous the population the larger the sample will be needed. The following factors need to be taken into consideration in making a decision as to the sample size:

- The degree of precision required between the sample population and the general population
- The variability factor of the population. (This is commonly expressed statistically as the standard deviation)
- The method of sampling employed (Leedy, 1974: 100).

However, a sample of 5 to 10% depending on town or village characteristics

is acceptable (Simpson-Hébert, 1983:19)

7. INTERVIEWERS

To a large extent, an interviewer influences the respondent's willingness and ability to respond to questions asked. Therefore, it is imperative that the interviewers be appropriately selected and trained to carry out the survey. Rural water supply and sanitation projects often involve small communities with dispersed settlements of approximately 200 households. In such cases then, it is ideal for the researcher to carry out the survey

himself or supervise interviewers closely in order to ensure reliability in data collection. If the researcher does not know the vernacular of that particular area, it is essential for him to employ the services of an interpreter. However, the interpreter should be made familiar with the objectives of the study as well as the questions before the actual interviews are conducted.

In big villages where it is not always possible for the researcher to undertake the study himself there is a need for selection and training of interviewers. Preferably, the interviewers must at least be trained in social science techniques in order to provide the planners with reliable information about the community. Therefore, professional elites like social workers, social anthropologists, nurses, teachers, agricultural extension officers and the like could be suitable for this kind of studies because of their knowledge and experience in interpersonal communication skills. In the absence of these professionals, then people who hold Junior or Senior Certificates could be considered for training. However, the training should be quite adequate in order to ascertain an understanding of research objectives and basic interviewing skills.

There is often a tendency by interviewers who are not properly trained to ask leading questions or even explain answers to respondents. Therefore, it is essential to ensure that interviewers will ask similar questions to all respondents. Other aspects that need consideration in selecting interviewers are responsibility, etiquette and commitment of interviewers to objectives of the study. The manner in which the interviewer presents himself and dresses will also have bearing on the success or failure of the interview.

8. ANALYSIS OF DATA

At the end of any survey, the raw data that has been collected need to be analysed and interpreted in such a way that everybody including a layman in the field of research could understand it. In order to arrive at a conclusion and make reasonable decisions, statistics is used as a tool in analysing the data. Statistics is defined as a language which, through its own special symbols and grammar, takes the numerical facts of life and translates them meaningfully (Leedy, 1974: 21). The observational type of studies in which the researcher describes what he has observed are called descriptive surveys. Therefore, the kind of statistics employed in purely descriptive studies often reveal the points of central tendency, the variability, and the degree of interrelationship between variables in the data, which will be discussed later. We will now turn to the discussion of measurement scales.

8.1 Measurement scales

The measurement of data is expressed by means of various scales of value. Four basic measurement scales are discussed:

Nominal scale

Nominal scale is the simplest level of measurement. The nominal data is being distinguished from all other data by simply assigning them a name. It merely expresses categorial classification without saying how much greater or better is one to another. Here numbers or symbols are merely used to identify the groups to which various objects or persons belong and these symbols constitute a nominal scale. For example, apples, toilets, rivers, boys, women.

Ordinal scale

The ordinal scale is a higher level of measurement than the nominal scale because it has been assigned an order of sequence. It indicates a measurement of the degree of difference. For example" Waterborne sewerage is better than the ventilated improved latrines or boreholes are better than wells. Another example of an ordinal scale is the system of grades within an organization - labourer, supervisor, manager, director and chief executive. Ordinal scales allow for classification of objects or people without saying exactly how much of the characteristics they possess except for the degree to which they possess the characteristics.

Interval scale

The interval scale is the next in refinement. It does not only rank objects with respect to degree to which they possess certain characteristics, but is also able to give the distances between them. It is, therefore, measured in terms of difference in standard units between one object and another. For example: Mary is taller than Tom, an IQ of ten points more than his brother's IQ.

However, the interval scale requires a physical unit of measurement which can be agreed upon and can be applied over and over again with the same results. Examples of these units of measurement are the metric system and the imperial system. Length, for example, is measured in metres or feet. There are no such units of intelligence which can be agreed upon. Given the unit of measurement, it is possible to say that the difference between the scores is twenty units, or that one difference is twice as large as a second (Hannagan, 1982: 71).

Ratio scale

The ratio scale is the highest level of measurement. With ratio scales the values are measured from absolute zero point and it is possible to indicate that one item is so many times as large as, as bright as, more powerful than another (Leedy, 1974: 115). For example: This water has twice as much chlorine as the other. A seventy year old lady is twice as old as a thirty-five year old person.

To conclude this session it is important to note that data have two other characteristics, namely, discrete and continuous. Discrete data are those which arise from the process of counting, for example the number of tables in a hall, the number of people in a factory and so on. Continuous data on the one hand arise from measurement process which is part of the continuum, such as height, time and age values. These data merge into an unbroken continuum of values (Leedy, 1974: 115).

Smit (1987: 188-189) makes the following important remarks about measurement scales:

- He states that qualitative variables are only measurable on the nominal or ordinal scale whereas quantitative variables are measurable on the ratio or interval scale and consequently on the ordinal and nominal scales.
- A statistical technique that has been developed for a variable that is measurable on a particular level of measurement is also applicable to a variable that is measurable on a higher level, although information is sacrificed on the latter case since complete use is made of the information that is present in the observations. Such loss of information generally leads to the use of a less effective statistical technique. The use of Spearman's coefficient of rank correlation (developed for ordinal data) in the case of observations measurable on the interval scale will for instance generally be less effective than the use of Pearson's coefficient of product moment rank correlation (developed for interval data) since the latter is permissible.

  Therefore, the nature and measurement scale of the variable under consideration, as well as the information that is known on the statistical distribution of the variable, are closely related to the statistical techniques which may or should be used in a particular case.

8.2 The role of statistics and useful techniques for descriptive surveys

The purpose of this discussion is merely to highlight some statistical techniques available for descriptive surveys. It is assumed that the researchers have a reasonable background of statistics. In no way, therefore, does the author pretend to offer an academic treatise of the subject matter. For more details it is recommended that statistical textbooks including the references given at the end of this report, be consulted.

As already mentioned, statistics takes numerical facts and translates them meaningfully. In that way, statistics assist us in narrowing the area of disagreement in order to arrive at a reasonable conclusion. The systematic collection of data distinguishes statistics from other kinds of information. Though already partly mentioned, it is worthwhile to repeat that statistics is divided into:

- Descriptive statistics which includes the presentation of data in tables as well as calculating percentages, averages, measures of dispersion and correlation in order to reduce it to manageable proportions.
- Inductive statistics which involves methods of inferring properties of a population on the basis of known sample results.

For our purpose, however, we will only deal with descriptive statistics which is relevant to the kind of studies we undertake.

8.2.1 *Measures of central tendency*

Measures of central tendency refer to any measure that indicates the centre of the distribution or the point around which the universe of data revolves. Averages are useful in summarising the data because they tell us something about an entire set of numbers. Therefore, averages provide descriptive features of one group relative to another. The most commonly used measures of central tendency are the arithmetic mean, the median and the mode. Each of these averages have their own unique characteristics which the researcher should have knowledge of in order to select the most appropriate one for particular sets of data. The three measures of central tendency are discussed thus:

The arithmetic mean

The arithmetic mean is the most commonly used measure of central tendency referred to mostly as an average in our daily life. It is defined as the sum of observations divided by the number of observations. The mean can be used for a set of numbers and for a frequency distribution. For example:

i. Four people earn R5, R7, R9 and R10 respectively. Their average earnings is calculated thus:

   Mean = Sum of observations _ Number of observations, or

   $x = \Sigma x/n$

   $= (R5 + R7 + R9 + R10)/4 = R31/4 = R7,75$

   Their average earnings is therefore R7,75 each. x signifies the mean, _x signifies the sum of money and n the total number of people involved.

   or

   four rural households consume the following litres per day 100, 80, 160 and 200 litres.

   Their average water consumption is

   $(100 + 80 + 160 + 200) \div 4$

   $\frac{540}{4}$

   $x = 135$ litres per day

ii. Arithmetic mean of a frequency distribution

   This can be described as the weighted arithmetic mean or arithmetic mean of grouped data. This is calculated by multiplying the item by the frequency, adding them up and dividing them by the sum of frequencies (Hannagan: 1987: 132). The formula is:

   $$x = \frac{\Sigma fx}{\Sigma f}$$

   Where Σ = the sum of
   f = the frequency
   x = the value of the items

With the use of the assumed mean this becomes

$$x = \frac{\Sigma fdx}{\Sigma f}$$

Where x = the assumed mean

d_x = deviation from the assumed mean

fd_x = the frequency times the deviation from the assumed mean

We shall illustrate the frequency distribution thus:

TABLE 1: THE ARITHMETIC MEAN OF A FREQUENCY DISTRIBUTION

PRICE (TO THE NEAREST R)	NUMBER OF PIT LATRINES	DEVIATION FROM ASSUMED MEAN	FREQUENCY X DEVIATION FROM ASSUMED MEAN
	(f)	(d_x) (x = 28)	(fd_x)
20	2	–8	–16
24	6	–4	–24
25	10	–3	–30
30	4	+2	+ 8
32	3	+4	+12
	25		-70+20
	$\Sigma f = 25$		$\Sigma fd_x = 50$

$$x = x \pm \frac{\Sigma fdx}{\Sigma f}$$

$$= 28 - \frac{50}{25}$$

$$= 28 - 2$$

$$= R28$$

The arithmetic mean is only used in interval and ratio levels of measurement. We will now discuss advantages and disadvantages of the mean (Hannagan, 1982: 137).

Advantages of the arithmetic mean:

- It is widely understood and the basic calculation is straightforward
- It makes use of all data in the group and it can be determined with mathematical precision
- It can be determined only when the total value and the number of items are known
- It is appropriate when data is symetric, that is as many observations below the mean as above.

Disadvantages of the arithmetic mean:

- A few items of a very high or very low value may make the mean appear unrepresentative of the distribution. It may be misleading when a few non typically large or small values are present and is sensitive to coding errors.

- When there are open-ended class intervals, assumptions have to be made which may not be accurate.

Although the arithmetic mean does take all the data and values into consideration, this major strength is also a weakness under certain circumstances. For example, the arithmetic mean of 20, 30, 20 and 200 is 68. This answer is not close to any of the actual values. However, because all the values are used including the extreme ones like 20 and 200 in the above example, it is the most preferred measure of central tendency than the others.

The median

The median is defined as the middle value in a set of numbers. If the number of values is odd, for example, 1,2,3 then the median will be 2. If the number of values is even, the median will be the mean of the two middle numbers, for example

1, 2, 3, 4 : $\frac{2+3}{2} = \frac{5}{2} = 2{,}5$ The formula is $M = \frac{n+1}{2}$

or

For example: The following numbers represents the distances people travel to the nearest borehole 1 km, 2 km, 6 km, 8 km, 9 km.

The median is 6 km.

The median is only used in ordinal, interval and ratio scales of measurement.

From a frequency table, the median class is the lowest class for which

the cumulative frequency exceeds N/2. The formula is

$$Md = \underline{l}t \left(\frac{N/2 - f}{f}\right) i$$

F = cumulative frequency corresponding to lower limit

f = number of cases in interval containing median

$\underline{l}$ = lower limit of interval containing median

i = width of interval containing median (Blalock, 1979: 65)

TABLE 2: MEDIAN WHEN DATA IS IN THE FORM OF A FREQUENCY DISTRIBUTION

Class interval	f	F	Number of cases less than
30 - 39	6	6	39
40 - 49	12	18	49
50 - 59	15	33	59
60 - 69	15	<u>48</u>	69
70 - 79	8	56	79
80 - 89	6	62	89
90 - 99	<u>4</u>	66	99
Total	66		

Formula: $Md = lt \left(\frac{N/d - f}{f}\right) i$

For example: $60 \left(\frac{66/2 - 48}{15}\right) 10$

$= 60 \left(\frac{33 - 48}{15}\right) 10$

$= 60(15/5)10$

$= 59$

Advantages and disadvantages of the median (Hannagan, 1982: 141):

i. Advantages of the median:

- Extreme high and low values do not distort it as a representative average. Therefore, it is useful for describing distributions in areas such as wages where a few extreme values would distort the arithmetic mean.
- It is straightforward to calculate even if not all values are known, or where there are irregular intervals.
- It is often an actual value and even when it is not, it may 'look' representative and realistic.

ii. Disadvantages of the median:

- It gives the value of only one item. The other items are important in ascertaining the position of the median, but their values do not influence the value of the median. If the values are spread erratically, the median may not be a very representative figure.
- In a continuous series, grouped in class intervals, the value of the median is only an estimate based on the assumption that the values of the items in a class are distributed evenly within the class.
- It cannot be used to estimate the sum of all the items.

The number of items multiplied by the median will not give the total for the data, therefore it is not suitable for further arithmetical calculations.

The mode

The mode is defined as the most frequently occurring value. This is an average that is frequently used in conversation when reference is made to such things as an 'average' income, an 'average' person and so on. It may be stated that 'the average family has two children', meaning that most families have two children. The arithmetic mean may show that the average family has 2,3 children, but although this may be mathematically correct, it may not appear sensible. In such cases as this, the mode may appear to be a more sensible average to use (Hannagan, 1982: 144).

For example: The mode of these figures 3, 5, 5, 2, is 5 because it is the most frequently occurring number. In a frequency distribution, the mode is the item with the highest frequency.

For example: The age group of children who fetch water is illustrated thus:

TABLE 3: FREQUENCY DISTRIBUTION OF THE MODE

Age group (years)	Number of children fetching water
14	3
16	15
18	25
20	20

The mode in this example is 18 because 25 is the highest frequency. It is stated (Hannagan, 1982: 145) that calculating the mode for a grouped frequency distribution is not easy because since a group frequency distribution does not have individual values, it is impossible to determine which value occurs most frequently. It is, however, possible to calculate the mode but it is not particularly useful. Therefore, for most purposes, the modal class is perfectly satisfactory as a form of description. For example:

TABLE 4: NUMBER OF PERSONS IN A HOUSE, 5 YEARS OLD, BUT LESS THAN 31 WHO ARE SUFFERING FROM TYPHOID

Age (years)	Number of persons
5 - 10	10
10 - 15	5
15 - 20	15
20 - 25	20
25 - 30	4
Total	54

The modal class in Table 4 is 20 because this class has the highest frequency. The modal class can also be illustrated by drawing a histogram. Examples of these will be given as appendices of this report.

To conclude this session, we will discuss the advantages and disadvantages of the mode (Hannagan, 1982: 147).

i. Advantages of the mode

- It is a commonly used average, although people do not always realize that they are using it.
- It can be the best representative of the typical item, because it is the value that occurs most frequently.

- It has practical uses. For instance, employers will often adopt the modal rates of pay, the rate paid by most other employers. Cars and clothes are made to modal sizes, houses are built on the basis of the modal average size of the family.
- The mode is often an actual value and therefore may appear to be realistic and sensible.

ii. Disadvantages of the mode

- It may not be well defined and can often be a matter of judgement.
- It does not include all the values in the distribution.
- It is not very useful if the distribution is widely dispersed.
- It is unsuitable for further or other kinds of calculation because of its lack of exactness.

Comment: Each of the measures measure a different property. All three should be reported and the differences between them can provide useful information regarding skewness, symetry etc.

8.2.2 *Measures of variability or dispersion*

Measures of central tendency are useful in summarizing the data by indicating the central point of the data in the distribution. Either than giving the location of the data, averages tell us nothing about how broadly the universe of data spreads. In view of the fact that distributions are not only clustered around the central point, it is important that we know something about the spread of the data in the distribution, hence, the need for measurement of dispersion. For example, the attitudes of household members towards their water quality is obviously different. Measures of central tendency will only indicate the typical attitudes of most people and tell us nothing about a few people who may have different attitudes. there are several measures of dispersion that include the range, the mean deviation, the variance and the standard deviation. However, the most commonly employed measure of dispersion is the standard deviation. We shall briefly discuss each:

The range

The range is the simplest measure of dispersion which describes the difference between the highest and the lowest values. The range can be defined as the highest value in a distribution minus the lowest (Hannagan, 1982: 155). For example the range of 90 and 70 is 20. The range is used in our everyday life in garages, supermarkets and so on to determine the price of goods. We often ask about price range of cars, carpets, blankets. The salesman will normally tell you that the price ranges between R50 and R150. This is the simple statement which provides the dispersion or spread of prices for certain goods. This indication of a spread is often misleading as it depends only upon two extremes. All measures of dispersion provide similar information to the range but expressed differently and with more precision.

The quartiles, deciles and percentiles

The median divides an ordered distribution into half, and in a similar way it is possible to divide distributions into quarters, tenths and so on. The most frequently used division is into quartiles. These divide an ordered distribution into four equal parts.

These values, denoted by Q_1, Q_2, and Q_3 are called the first, second and third quartiles respectively, the value Q_2 being equal to the median.

Similarly the values which divide the data into ten equal parts are called deciles and are denoted by D_1, $D_2 ..., D_9$, while the values dividing the data into one hundred equal parts are called percentiles and are denoted by P_1, $P_2 ... P_{99}$. The 5th decide and the 50th percentile correspond to the first and third quartiles respectively.

For example: In a distribution of 100 items the quartiles will be the values of the 25th (Q_1) and the 75th (Q_3) items. With the median, which will be the value of the 50th item, the two quartiles will divide this distribution into four equal parts: 1 - 24, 26 - 49, 51 - 74, 76 - 99.

The interquartile range

Hannagan (1982:156) defines the interquartile range as the difference between the upper quartile and the lower quartile (Q_3 - Q_1). Table 2 shows the wage list of a company employing twenty people. The wages of employees 5, 10 and 15 on this list divide the distribution into four equal parts. (1-5, 6-10, 11-15, 16-20).

The position and value of the quartiles are:

$$Q_1 = \frac{n}{4} = \frac{20}{4} = 5$$

This is the position of the lower quartile and the wage of the fifth employee is the lower quartile wage: R50.

$$Q_3 = \frac{3n}{4} = \frac{60}{4} = 15$$

TABLE 5: THE INTERQUARTILE RANGE

EMPLOYEES	WEEKLY WAGE (R)	EMPLOYEES	WEEKLY WAGE (R)
1	40	11	89
2	42	12	93
3	43	13	97
4	48	14	100
5	50	15	110
6	60	16	140
7	62	17	200
8	65	18	210
9	71	19	212
10	80	20	220

This is the position of the upper quartile and the wage of the fifteenth employee is the upper quartile wage: R110.

The interquartile range $= Q_3 - Q_1$

R110 - R50 = R60

It is also possible to arrive at the semi-quartile range by dividing the interquartile range by two:

The semi-interquartile range = $\frac{Q_3 - Q_1}{2}$

$$\frac{R110 - R50}{2} = \frac{R60}{2} = R30$$

The interquartile range helps to summarize and clarify the distribution. In Table 1 the interquartile range (R110 minus R50) covers those employees earning the middle range of wages, and this includes 50% of the employees working in the company. The interquartile range is less dependent on extreme values as illustrated above.

The mean deviation

The mean deviation is defined as the arithmetic mean of absolute differences of each score from the mean (Blalock, 1979: 78). The mean can be used to determine the deviation of each score from the average because it is the most satisfactory single measure under most circumstances. The advantages of the mean deviation are that it includes all observations and it is also easy to define. The mean deviation is calculated thus:

Formula: Mean deviation = $\Sigma \left| \frac{xi - x}{N} \right|$

For example: The mean deviation of numbers 11, 14, 17, 10 and 8 = 12.
In subtracting 12 from each of these numbers, adding the results and deviding by 5 we get

$$\frac{1 + 2 + 5 + 2 + 4}{5} = \frac{14}{5} = 2{,}8$$

We may, therefore, say that on average the scores differ from the mean by 2,8.

The variance and standard deviation

The variance is defined as the square of the standard deviation and is a widely used measure of dispersion in statistical analysis. The standard deviation which is a square root of the variance on the one hand, is defined as a measure of dispersion which uses all the values in a distribution in the sense that every value contributes to the final result in the same way that every value contributes to the calculation of the arithmetic mean (Hannagan, 1982: 159). The standard deviation is the 'standard' measure of dispersion because it is very useful both practically and mathematically.

The standard deviation is derived from the variance and, therefore, has the same units as original observations. This makes it possible to compare the standard deviation with the mean, both being expressed in the same units. The standard deviation and the variance are used because of their statistical properties and are more useful measures of dispersion. The total population has a mean and a standard deviation which will describe the position and dispersion of its frequency distribution. Hannagan (1982: 160) states that the standard deviation shows the dispersion of values around the mean. The greater the dispersion, the larger the standard deviation.

To calculate the standard deviation we first calculate the variance which is abbreviated as S^2, from which the standard deviation (S) may be found by taking the square root. Therefore, the standard deviation is the square root of the variance.

Formula $V = \frac{\Sigma(xi - x)^2}{N - 1}$

$i = 1$

For example: $x_1 = 4$

$x_2 = 6$

$x_3 = 8$

Mean: $x = \frac{4 + 6 + 8}{3}$

Therefore $v^2 = \frac{(x_1 - x)^2 + (x_2 - x)^2 + (x_3 - x)^2}{N - 1}$

$$= \frac{(4 - 6)^2 + (6 - 6)^2 + (8 - 6)^2}{2}$$

$$= \frac{4 + 0 + 4}{5}$$

$$= \frac{8}{2}$$

$$= 4$$

Variance is 4

For a standard deviation we, therefore, take the square root of the variance. Taking the variance of the above example, we work out the square root of 4.

The standard deviation is, therefore $\sqrt{4}=2$

To conclude this section, we may say that the measures of dispersion, as much as the measures of central tendency are helpful in describing the distribution because this indicates whether or not the data is clustered close together or is well spread out. For instance, economists may be as interested in changes in the distribution of incomes as in changes in the averages (Hannagan, 1982: 170). For most of the water supply and sanitation studies, however, we tend to employ the measures of central tendency than the measures of dispersion. The standard deviation would be useful in finding out, for instance, the spread of household incomes.

8.2.3 *Correlation*

To this point we have been discussing the central point around which the data revolves and the extent to which the data is spread. Life situations cannot be adequately described by a static system where only central balance points and amplitude of dispersion account for the world around us. In the real world, facts of life exist in dynamic relationship with each other (Leedy, 1974: 28). therefore, correlation is concerned with whether or not there is any linear association between two variables. If two variables are related to any extent, then changes in the value of one are associated with changes in the value of the other (Hannagan, 1982: 198).

For example, we would like to know if there is any relationship between the existence of waterborne diseases and the infant mortality rate. It is useful to establish this relationship as it will help to plan aspects that need consideration. For instance, if it is established that there is any relationship between waterborne diseases and infant mortality rate, it is easy to convince health and water authorities to improve the water quality in a particular area. Plans and predictions based on evidence rather than speculations are worthwhile in making decisions. Correlation may provide insight into the cause-effect relationship of certain aspects in life.

The correlation coefficient can range from -1 to +1. If it is -1, correlation is perfect negative and if it is +1, it means perfect positive correlation. Zero (0) means that there is no correlation. When correlation is perfect, all points will lie on a single straight line, otherwise in the case of imperfect correlation, points will be scattered around (see Figures 5 to 9).

There are several appropriate correlational techniques for different sets of variables such as product-moment correlation, rank difference correlation and so on, however, we will not discuss them in this report. To conclude we may say that one variable correlates meaningfully with another only when there is a common, causal bond that links the phenomena of both variables within the framework of common relationship (Leedy, 1974: 31). For example, high qualifications usually correlates with high income.

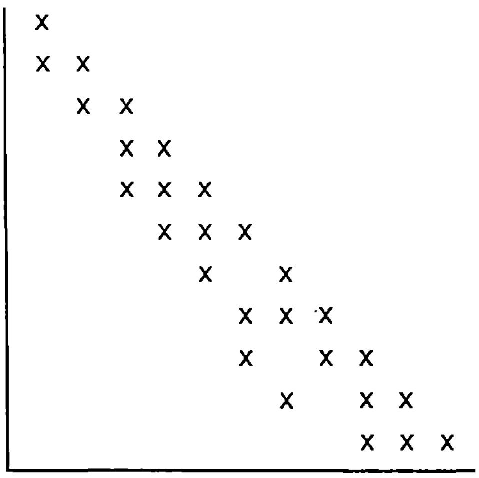


Figure 5: *Negative correlation*

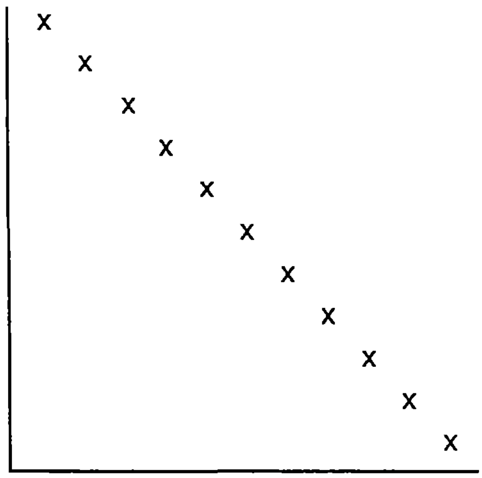


Figure 6: *Perfect negative correlation*

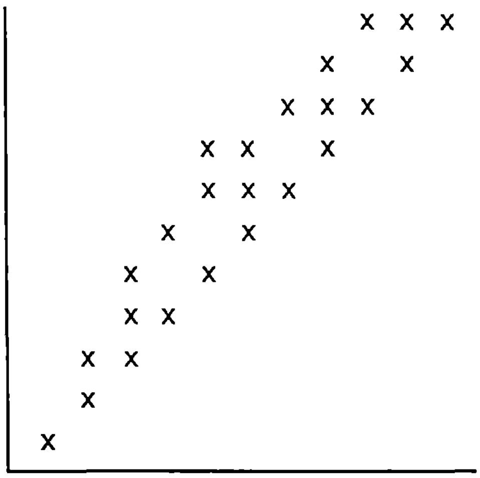


Figure 7: *Positive correlation*

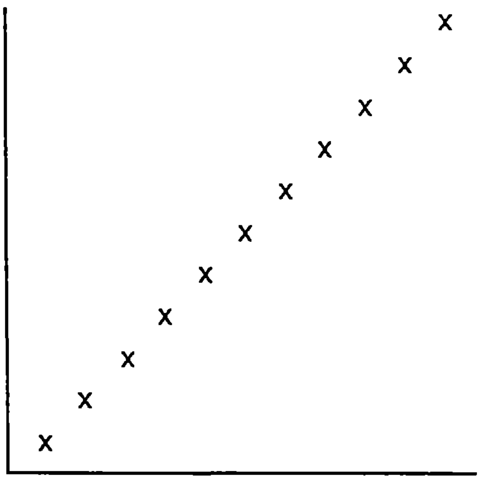


Figure 8: *Perfect positive correlation*

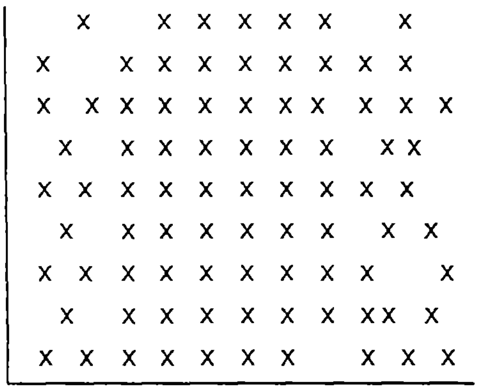


Figure 9: *No correlation*

8.3 Computer programmes and data analysis

It is well known that the usefulness of any survey is inversely proportional to the time before the results are made known. There are computer packages available to perform even the most complex of statistical calculations. It is, therefore, recommended that use be made of these packages to speed up the data processing. However, it should be remembered that a computer cannot do research. It is a machine that cannot think, thus its effectiveness depends on the operator who is a human being.

8.4 Facet analysis of data

Facet analysis is a computer programme which can be used as a supplement to social science data analysis. Facets are categories of observations within which there are mutually exclusive elements, for example "sex" could be a facet with 'male' and 'female' as its elements. This technique treats interrelationships in a geometrical way and displays them in a visual form, and this facilitates appreciation of data especially by people without statistical or mathematical backgrounds. It is designed to determine the structure of intercorrelations between variables.

Facet analysis is based on the same body of mathematical knowledge as the rest of non-parametrical statistics. Therefore, this technique cannot replace but supplement conventional statistics. The usefulness of this technique as a supplement to conventional statistics has been demonstrated in an evaluation study of water supply and sanitation projects in the Valley of a Thousand Hills, near Durban (Mogane 1987: Internal report). This technique showed graphically the groupings of people according to their different perceptions and attitudes towards their water supplies and sanitation. Therefore, facet analysis as a supplementary means of sociological data analysis is considered well worth pursuing to facilitate a better understanding of social science research data.

9. CONCLUSION

The current development strategies strongly advocate community participation and involvement in projects that are aimed at improving people's lives. It is a recognized fact worldwide that projects that take people into consideration tend to be more successful than those that do not. In order to involve people in any project, the planners need to be socially informed about aspects of community life. Of course there are various methods of gathering social science data which have not been discussed in this report such as postal questionnaires. These methods were deliberately discounted due to their impractical nature for rural settings. For example, it would be difficult, if not impossible for rural communities, most of whom are illiterate to complete question- naires independently.

Therefore, methods discussed in this report are qualitative in nature as they require personal contact between the project team and intended beneficiaries. These qualitative techniques offer a golden opportunity for both the community and project team to know each other's aspirations and needs. Moreover, qualitative techniques are essential in determining the long term working relationship with the community once the project commences.

10. ACKNOWLEDGEMENTS

I gratefully acknowledge the generous courage, support and motivation of all my colleagues, in particular Mr Ian Pearson in writing this report. I am also indebted to Ms Dudu Radebe, a colleague from the Division of Building Technology, for her comments on the draft document and Messrs B J Dixon of the Opinion Survey, Human Sciences Research Council and M A Trollip of the Centre for Advanced Computting and Decision Support, CSIR for their edition and constructive criticism of this report.

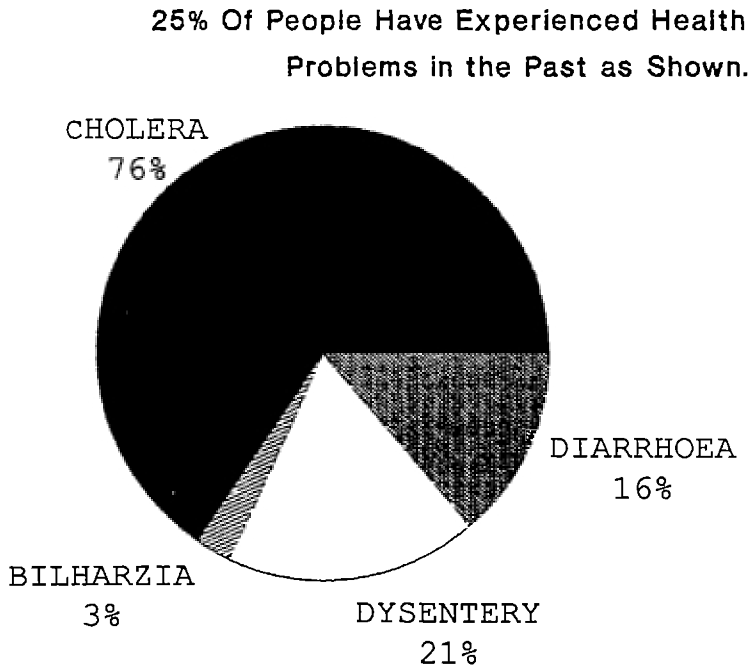
Despite borrowings from other sources, the information reflected herein represents the views of the author and not necessarily those of her employer, CSIR.

11. REFERENCES

1. BLALOCK, H M. 1972. Social statistics, New York: McGraw Hill
2. CERNEA, M M. 1985. Putting people first. Sociological variables in rural development. Washington D C: The World Bank
3. CHANDLER, M M. 1986. Achieving success in community water supply and sanitation projects. *Waterlines,* vol. 5, no. 2, October. London: Intermediate Technology Publications
4. ELMENDORF, M and BUCKLES, P. 1980. Socio-cultural aspects of water supply and excreta disposal. Transportation, Water and Telecommunications Department: The World Bank
5. HANNAGAN, T J. 1982. Mastering statistics. Britain/Hong Kong: MacMillan Press
6. LEEDY, P D. 1974. Practical research - Planning and Design. New York: MacMillan Publishing Co.
7. MILLS, J J. 1987. A sociological study of water and sanitation related diseases as a socio-political phenomenon in South Africa: Submitted in accordance with the requirements for the degree D.Litt, et Phil. Pretoria: University of South Africa
8. MOGANE, S B. 1987. A sociological evaluation of village water supply project in the Valley of a Thousand Hills. Internal Report: CSIR, Pretoria (available on request)
9. MOGANE, S B. 1987. Facet analysis as a supplementary means of social science data analysis (with reference to rural water supplies in the Valley of a Thousand Hills, KwaZulu). Internal Report: CSIR, Pretoria (available on request)
10. RAYMOND, L G. 1980. Interviewing: Strategy, Techniques and Tactics. USA: Dorsay Press
11. SIMPSON-HéBERT, M. 1983. Methods for gathering socio-cultural data for water supply and sanitation projects. Washington D C: The World Bank - Technology Advisory Group.
12. SMIT, P C. 1987. Survey methods and practice. Human Sciences Research Council report: Pretoria

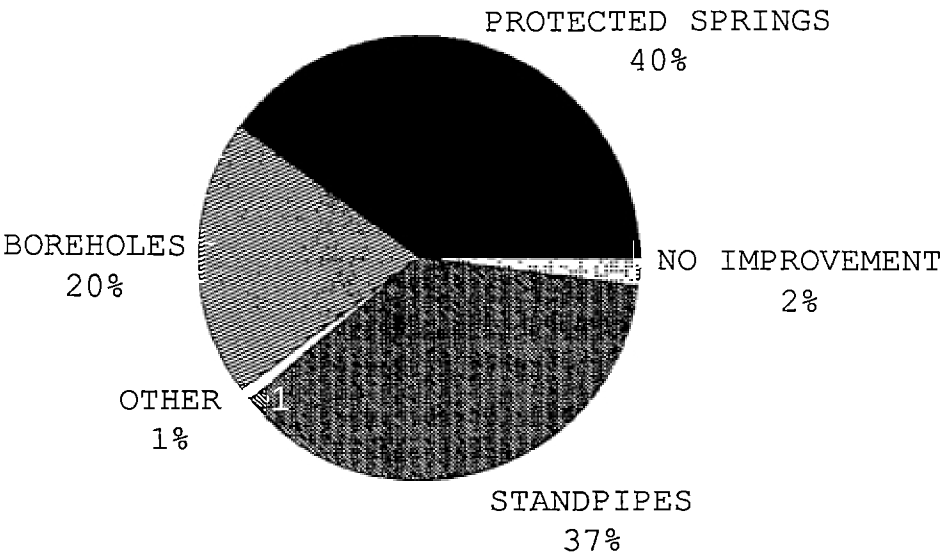
Summary: Results of the feasibility study of water supply and sanitation project in the KwaHlophe village of Ndwedwe, KwaZulu.

NDWEDWE SPRING PROTECTION SURVEY



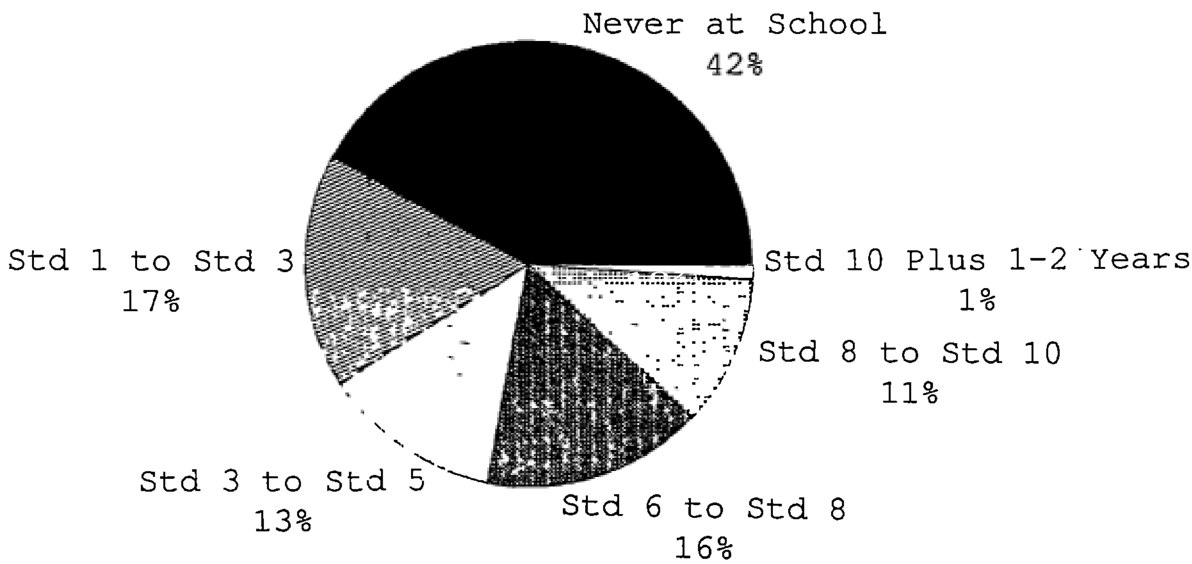
Respondents who experienced health problems in the past

NDWEDWE SPRING PROTECTION SURVEY
The Improvement People Would Like to See



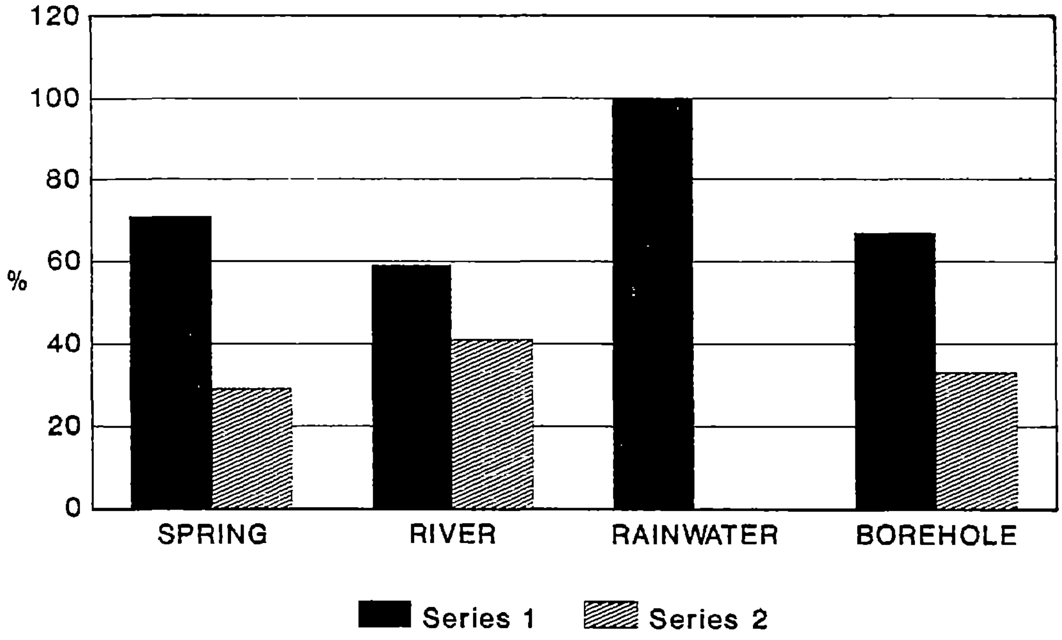
Water supply improvements the respondent would like to see

NDWEDWE SPRING PROTECTION SURVEY
Respondent Education Level



The educational level of respondents

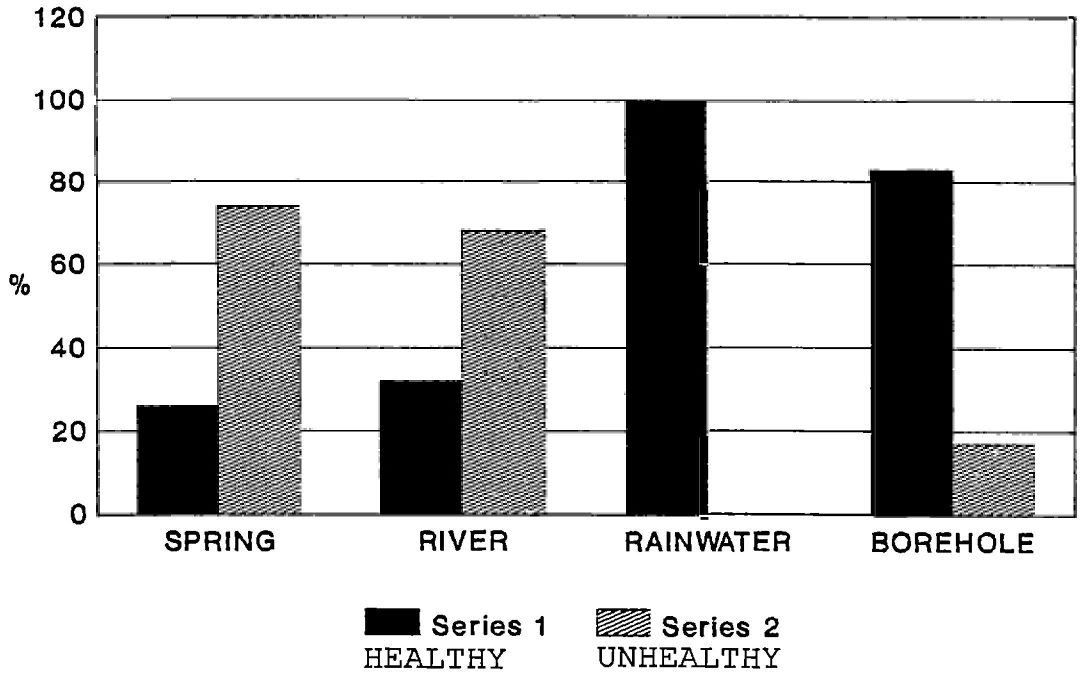
NDWEDWE SPRING PROTECTION SURVEY
Relation Between Source & Taste



Series 1 - Like   Series 2 - Dislike

Relationship between the source and taste

NDWEDWE SPRING PROTECTION SURVEY
How Do People Perceive Their Water ?



People's perceptions of their water

WATER CARE PROGRAMME

DIVISION OF WATER TECHNOLOGY
CSIR
Box 395
Pretoria
0001

Tel : (012)841-2272 or 841-2254
Fax : (012)841-4785
Telex : 3-21312 SA

The Appropriate Technology Group of the Water Care programme produces technical guides and videos on water supplies for small communities which are available on request.

Presently available are:

- Healthy water for your family
- Water filter for your home
- How to protect a natural spring
- How to build a small ferro-cement water tank
- Rainwater harvesting