INTRODUCTORY RESEARCH METHODOLOGY:
EAST AFRICAN APPLICATIONS

By

Kenneth Prewitt

OCCASIONAL PAPER NO. 10

INSTITUTE FOR DEVELOPMENT STUDIES
UNIVERSITY OF NAIROBI
P.O. Box 30197
Nairobi, Kenya

JANUARY, 1975

INTRODUCTORY RESEARCH METHODOLOGY:
EAST AFRICAN APPLICATIONS

By

Kenneth Prewitt

## ABSTRACT

This is a textbook in introductory social science methodology, with an emphasis on East African illustrations and applications. It covers standard topics in social data collection: sampling theory, survey questionnaires and interviewing, use of public documents and aggregate statistics, and systematic observation. It then reviews basic concepts in statistical data analysis, including simple correlations and tests of statistical significance.

The text is written for beginning students in the social sciences and for other persons, especially civil servants, whose occupation requires that they be knowledgeable consumers of social science research reports.

The emphasis throughout is placed on conceptual understanding rather than on computational knowledge. The text is not a 'how to' book in that it does not instruct the reader how to become a practicing social scientist. Rather it tries to tell the reader what the practicing social scientist is doing, and why. The purpose is to help the reader become an intelligent user, and critic, of social science research.

INTRODUCTORY RESEARCH METHODOLOGY:
EAST AFRICAN APPLICATIONS

Kenneth Prewitt

Institute for Development Studies, University of Nairobi

Abstract

Author's Forward

This text was written to aid in the teaching and learning of social science methodology in East African universities, research institutes, and relevant departments in the civil service.

The author considers the text to be unfinished, and invites readers who wish to contribute their own experiences and criticisms toward the improvement of the text. The text is unfinished in several respects. Perhaps most important, the implicit pedogogy adopted in writing the text has not been tested. There are undoubtedly ideas and concepts which are poorly expressed, and thus will be difficult to teach and to learn. Use will reveal these flaws.

Moreover, the text is unfinished in what it includes. Any textbook writer faces difficult choices of inclusion and exclusion. Whether the correct choices have been made is finally a judgment to be made by the readers. If readers wish to petition for the elaboration of certain issues or the inclusion of new topics, there will be an effort to accomodate them in subsequent editions.

Although the text is not intended as a survey of social science research in **East Africa,** it does attempt to use East African materials to illustrate basic issues in research methodology. No doubt there are other illustrative materials available which could elaborate on issues now discussed, and improve on the presentation of the issues. The present illustrations draw heavily on Kenyan research, largely because the author was based in Kenya while writing the text. A subsequent edition of the text would be improved by the inclusion of more extensive illustrative material; and readers who wish to do so are invited to supply examples of research techniques and applications from their own experiences.

Ms. Sidney Westley, Publications Editor at the Institute for Development Studies, University of Nairobi, Kenya, has performed invaluable services in preparing the manuscript and in shepherding it through the production process. Criticisms, comments, or relevant material which might improve any subsequent editions should be sent to her at I.D.S. In preparing this text the facilities of I.D.S. were made available to the author, for which he thanks Dr. Mugo Gachuhi, Acting Director, and Ms. Mary Kempe, I.D.S. Research Documentalist. A grant from the Rockefeller Foundation provided the time necessary to work on the text.

CHAPTER 1

SOCIAL SCIENCE RESEARCH METHODOLOGY

What Is Research Methodology?

Research methodology can be simply defined as:

> Systematic research procedures and techniques which
> help the researcher to avoid self-deception.

Why a definition which stresses 'self-deception'? Because we assume
that the research scientist wants to accurately describe and explain the thing
he has observed. If we read in a research report that there is no relation-
ship between the income of a farmer and whether he practices multiple-cropping,
we assume that the researcher honestly believes there to be no relationship.
Yet a researcher can be honest but wrong. This is where methodology comes
in. The trained researcher uses procedures which minimise the probability
that he is inaccurate. He especially wants to guard against being unwittingly
inaccurate, that is, to guard against self-deception. For if he deceives
himself, he necessarily risks deceiving those who rely on his findings for
action or edification.

The 'truth-value' of empirical statements is a very complex topic;
it has been debated for centuries by philosophers of science (and today is
the branch of philosophy known as epistemology). We cannot deal adequately
with the philosophical issues raised. What we can do is to describe some of
the practical ways in which the researcher tries to increase the 'truth-value'
of his statements, another way of saying that he tries to avoid self-deception.

To say that methodology is the systematic attempt to avoid self-
deception only states a principle. It does not yet tell us what methodology
is. This entire text is about methodology, and thus the real answer to the
question comes only at the conclusion. But we begin to understand by
contrasting what the untrained social observer might do with what the
**trained** social observer would do.

Consider a government official who is sincere in trying to under-
stand a particular social situation, but who is not trained in the procedures
of research.  The official is in the Ministry of Agriculture and is responsible
for the effectiveness of the agricultural extension services.  One weekend
he drives into the countryside to do a spot check on the work of his extension
officers.  On this trip he notices a farmer spraying his coffee plants, so
the official stops the car and asks the farmer what insecticide he is using.
The official is pleased to learn that the insecticide is the variety recently
recommended by the Ministry to the staff of the extension service, and he
concludes that the extension worker in this area is doing his job.

Consider the possible sources of error (and thereby self-deception).
There is first the problem of the sample: can we be certain that the farmer
questioned is typical of other farmers in the area? Perhaps he is the only
farmer who happens to be using that insecticide, and it was just an accident
that the ministry official stopped at this farm rather than any one of dozens
of farms where the insecticide was not being used.  The trained research methodo-
logist would be sensitive to the possibility that this was an accidental
occurrence.  He would understand the principles of sampling; and he would
know that there are specialised procedures for determining the likelihood
that this particular farmer is representative of other farmers.  That is, the
trained observer applies procedures to reduce self-deception, the self-
deception which occurs when something atypical is thought to be typical.

Another source of error is the data collection instrument.  The
official who questioned the farmer assumed that the farmer accurately reported
on the type of insecticide being used.  But this may not have been the case.
The farmer may not even know the name of the insecticide, and thus simply
mentioned the first thing which came to mind.  And the first thing which came
to mind happened to be a type he had heard discussed at the cooperative last
week, though he had not bothered to get some for himself.  The trained
observer would check on the answer given by the farmer, perhaps by asking
where and when he bought the insecticide, how much it cost, what kind of
package it came in, and so forth.

We know as well that persons being interviewed often say what they think the interviewer wants to hear. If the farmer saw that the official was being driven around in a Ministry of Agriculture vehicle, he might report using the up-to-date insecticide when in fact he was using a cheaper and less reliable type. It is not that the farmer deliberately wants to mislead. It is just that he wishes to please. And in this he is successful, though at the cost of providing **inacurrate information.** Unfortunately the official, not knowing this, is deceived. A trained observer is less likely to make this mistake. He knows that data collection instruments of all kinds, and especially interviews or questionnaires, can introduce errors and bias. The trained observer uses procedures which attempt to correct for these errors.

The ministry official opens himself to another major source of error. He makes a dubious inference. Because he learns (or thinks he learns) that the farmer is using the recommended insecticide, he infers that this is the result of effective extension work. Yet there is nothing in the information he is provided which warrants this inference. The farmer may be typical (no sampling error) and may in fact be using the recommended insecticide (no data collection error) but not have learned about the insecticide through channels having anything whatsoever to do with the extension service. Here would be a case where the data are accurate but the conclusions drawn from the data are inaccurate. And again the untrained observer **has** let himself be deceived.

There are many statistical procedures which protect the trained observer against faulty inferences, or at least lessen the chances of making them. There are ways to determine whether the rate of adoption of a new insecticide is positively related to the efforts of extension workers. There are even ways to predict just how much the rate of adoption might increase depending on how much more effort is put in by the extension workers. And there are ways to calculate how likely is the relationship discovered to be the 'true' relationship.

The research procedures used by the trained social scientist do more than protect him against inaccurate conclusions, though this is a major purpose of methodology. Two further aspects of research methodology are important: data organisation and propositional thinking.

The social world is enormous and complex, and it follows that the data available to the social scientist are enormous and complex. The simple observation that the farmers of Kisii are shifting to small-holder tea farms immediately suggests to the rural development scholar a host of important research questions, all of which would require considerable data to answer:

> What kinds of farmers are shifting? What kinds of farmers are not shifting? What are the marketing arrangements? What has the role of the government been? What are the implications for the economic growth of Kisii? What are the implications for politics in Kisii? How does the shift affect the relationship between Kisii and the rest of the country? What is being taken out of production? Are there implications for family structure? Might health or educational services be affected by changes in the economic basis of the District?

No social scientist is likely to consider all of these research questions simultaneously -- though a major study of integrated rural development might. Yet whatever limits are placed on a study of tea-growing in Kisii, the range of data available to the researcher will be considerable. And the data will have to be effectively organised. This organisation of data is facilitated if appropriate research methods are used. Research methods classify (farmers who have shifted and farmers who have not); they count (What proportion of land has gone into small-holder tea?); they relate (Are tea-growing farmers closer to market roads?).

The second purpose of research methodology is to assist the social scientist to think propositionally. The best translation of this idea is captured in the oft-used phrase 'If...then...statements'. If a farmer is close to market roads, then he will be more likely to shift to tea-growing. If a whole farming area changes to a new crop, then there will be a shift in the persons who control local political and economic institutions.

This 'if...then...' thinking is the stock in trade of the social scientist. It is a different way of saying that he is searching for the social conditions and causes which bring about patterns of social behaviour. The social scientist is always asking, 'How did this particular social pattern or arrangement come about?' And he is asking, 'What factors or events might change the present social pattern, and change it in what direction?' Of course

such questions are asked by persons other that social scientists, but the
training of the social scientist leads him to ask the questions in such a
way that they can be answered with systematic data.

## Social Science Data

We have talked about research methodology without saying much
about social science data. Simply, social science data means information
or evidence about people. The social scientist wants to understand how
people earn their living, how they raise their children, how they change
from one way of doing things to another way, how they establish and use
technologies, how they play and how they worship, how they behave socially
and politically, and how they deal with each other.

The social scientist has devised many different ways for getting
the information he wants. Sometimes (in fact, very often) he simply asks
people questions. At other times he stands aside and observes how people
behave in different situations. When questioning or observation is
impractical the social scientist might turn to the records which people
leave of their behaviour; the political scientist studying voting behaviour
by collecting election statistics or the anthropologist studying work habits
by analysing the tools people use.

It is important to realise just how extensive is the range of data
available. Try a simple exercise. Assume a researcher is asked to organise
a study of citizenship training in Tanzania and Kenya. He is supposed to
determine which country's school system is doing the most effective job of
citizenship instruction.

What type of data would he collect? It turns out that many types
of data are available. He could interview pupils directly. Then he might
want to compare their citizenship values and their political information with
what government leaders expect pupils to believe and know. The second part
of the task would involve analysis of government publications or speeches by
government leaders. A very different research strategy would involve study
of citizenship materials used in the schools; which country's school system
has the more extensive curriculum materials for citizenship instruction.

Another approach is to interview teachers or headmasters to see how committed they are to citizenship teaching. But perhaps he considers none of these research strategies suitable, and instead thinks that the critical citizenship training takes place outside rather than inside the school room. Starting from this premise he might choose to observe how pupils behave on the playground or when working on the school farm or when engaged in school club activities.

The choice of research strategy and thus the choice of data collected depend on several factors. Most important is a clear understanding of the purposes to which the data are to be put. If 'effective citizenship training' means how strongly pupils believe in national policy, then some method of directly questionning or observing students will be necessary. If 'effective citizenship training' means the commitment with which educational authorities accept this responsibility, then needed is information about the resources given to this part of the school curriculum.

Cost is another factor in deciding what sort of data to collect, second in importance only to making certain that the data match the research purposes. For instance, a study of the textbooks and curriculum materials used in Kenyan and Tanzanian schools is much less costly than a survey of student political values, but it would not be a study design which could tell one whether the materials were actually being learned. A survey is often very costly, but has the advantage of allowing the research to cover a wide variety of schools in both countries. It also has the disadvantage of providing superficial data unless designed very carefully. Direct observation can provide in-depth material, is relatively cheap in terms of materials, but expensive in terms of research time. For instance, it would be difficult to observe in depth more than one or two schools in each country and thus one immediately confronts the problem of whether the schools chosen are truly representative.

Related to costs are also a few practical considerations. Sometimes the data wanted are not available to the researcher. This often happens to the political scientist because he finds that important government activities take place in secrecy. The political scientist might learn that the issue of

'citizenship training' was on the agenda of a high-level meeting in the
Ministry of Education, but be unable to obtain a copy of the minutes of that
meeting.  Another practical consideration is how often the thing one wants
to study happens.  For example, a researcher might wish to observe how
schools **celebrate national independence day**, but find that because the
celebrations are going on simultaneously in all the schools he can observe
only one or two cases.  There is also the problem of whether the researcher
can effectively plan  his research.  Perhaps he thinks it would be useful to
see how school authorities handle school strikes.  But how can he be at the
right school at the right time since school  strikes are not normally
announced in advance (and talking with authorities and students after the
event can produce distorted reports).

Thus we see that all data sources are not available at all times
to all researchers.  Research purposes set limits to the type of data one wants
to collect, but so also do research costs and related practical considerations.

## Reactive and Non-Reactive Research

Social research involves a relationship between the researcher and
the respondent, the respondent being the person about whom the researcher
wants information.  Note that we say 'about whom' and not 'from whom'.  In
this choice of words we have already encountered an important principle.  The
researcher can get information about an individual (and therefore about a
collection or sample of individuals) without getting that information from
the individual.  That is, social research can be either reactive or non-
reactive.

Reactive Research:  In this type of research, the respondent is presented
with a stimulus by the researcher.  The best and most frequent example is the
survey questionnaire.  The questions asked in the interview are stimuli, and
the respondent responds (you can see why the term respondent is used in
survey research).  Questionnaire research is not the only kind of reactive
research.  Psychologists, for instance, sometimes present respondents with
non-directive stimuli such as ink blots and ask the respondent to describe
what he sees.

<u>Non-reactive Research</u>:  In this type of research the information about the respondent is gathered without direct interaction between researcher and respondent.  The person being studied is not asked to react to a stimulus presented by the research situation.  Perhaps the best example is research based on public record data.  Political scientists study policy priorities by collecting budget statistics; sociologists study geographical mobility through census data; anthropologists study beliefs by analysing the religious artifacts of a people.

Observational research is generally a mixture of reactive and non-reactive techniques.  The researcher can observe behaviour without presenting a stimulus, as when the child psychologist observes child-rearing practices.  But usually the observer will also ask questions in order to fully understand the social meaning of the behaviour he observes.

The primary advantage of reactive techniques is the control that the researcher has over the stimulus he presents, and thus his ability to get the information he wants.  But this is also partly a disadvantage, for in presenting a stimulus the researcher becomes part of the social situation and thus **risks** altering the normal behaviour of the people he is **studying.**

The primary advantage of non-reactive techniques is that the researcher stays out of the research situation, and thus is less of a distracting or biasing factor.  But to rely only on non-reactive techniques is to be dependent on data collected by someone else.  And such data may only partially satisfy the research needs.  If, for example, the government publishes election statistics for national but not local elections, the political scientist using public records to study voting behaviour has no way of knowing whether participation differs from one type of election to another.

The distinction between reactive and non-reactive **research** techniques is useful, but should be applied cautiously.  Very many studies simultaneously employ more than one research technique.  And in a 'multi-method study' the re-searcher often chooses some methods which are reactive and other methods which are non-reactive.  This has particularly been the practice in East Africa, where observation and informal interviewing have been used extensively to supplement standard survey techniques.

Quantifiable Data

Although social science depends on many different types of data
collected in many different ways, the research methods described in this
text have one very significant limitation.  They are limited to what are
usually called 'quantifiable data'.  When a particular piece of information
is collected in a standardised **form** from a large number of units we refer to
the result as quantifiable data.

All three parts of this definition are important.  First, a
'particular piece of information' means something like household size or
level of political participation or use of mechanisation in farming or
attitude toward birth control.  The information must be collected from each
unit in the study: the size of every household in the study must be recorded;
the participation rate of every voter in the study must be measured; the
extent of mechanisation on every farm in the study must be determined; the
attitude toward birth control of every parent in the study must be ascertained.

Secondly, for the data to be quantifiable the piece of information
must be collected in a 'standardised form' from each unit in the study.
Should a household include married males who work in the city and visit the
rural household only on weekends?  Until the researcher answers this question,
there can be no standardised measure of household size.  And when he has decided
how to measure household size, the measure adopted must be applied uniformly
across all households in the study.  The same need for standardisation applies
to measures of political participation or farm mechanisation or attitude
toward birth control.  Social data are not quantifiable unless they are
collected in a uniform manner from every unit in the study.

The third part of the definition refers to a 'large number of units'.
Just what number constitutes a 'large number' cannot be definitely stated; it
will depend upon the research goals and upon the type of units being studied.
In Kenya there cannot be a comparative study of social services in different
provinces in which the number of units is larger than seven, because there
are only seven provinces.  Normally, however, there are many more units
available.  This is always so when the units studied are individuals or
individual households.  As a practical matter, many of the statistics introduced

in subsequent chapters of the text do not work well unless there is a minimum of fifty or so cases in the study (though some of the statistics can be adjusted when the sample is smaller than this).

From the definition of quantifiable data it is clear that the social scientist is involved in measurement, that is in assigning numbers to indicate how much of some characteristic each unit in the study has. The size of the household is measured by counting the number of people who live there; the income of a household is measured by adding the earnings of all the household members.

But even in these comparatively straightforward examples we see that measurement is no simple task. It is even more difficult when the measurements have to be created out of the research procedures themselves, what can be called 'artificial measures'. For instance, the researcher **might** be able to ask the head of the household how large the household is, but he cannot ask the respondent what his score is on a 'farming innovation scale'. This latter measure must be created out of different questions which only indirectly measure what the researcher is interested in.

Social scientists operate with a handicap that should be bluntly acknowledged: social science has no standardised measuring instruments which can be universally or even widely applied. It is in this regard that the social scientist is at a tremendous disadvantage, a disadvantage not felt by his colleagues in the natural sciences. The natural scientist has scales which always measure weight in the same way, thermometers which always measure temperature in the same way, gauges which always measure pressure in the same way, rulers which always measure length in the same way.

No comparable measuring instruments exist in the social sciences. We cannot say, "Here is a measuring instrument which identifies levels of rural development. Apply it in the Gulu District of Uganda and report the results." Nor can we say, "Here is a measuring instrument which determines the level of domestic political violence. Use it in Ethiopia to see what level of violence is associated with a major transformation of the political structures." In fact, we cannot even say, "Here is a measuring instrument

for calculating the size of rural households. Use it in a comparative study
of Nigeria and Zambia to see the relationship between household size and
rural to urban migration."

Faced with the complexities and ambiguities of measurement, one might
conclude that social science has no business attempting to base conclusions
on quantifiable data. There are persons who do take this position, and among
them are some talented social scientists who rely only on non-quantifiable
data. The skeptic's position is a reasonable one, but should be considered
against several observations.

Compared with other sciences, the social sciences have a very short
history. The use of systematic quantifiable social data, with a few exceptions
such as Durkheim, dates back only a few decades. Survey techniques, for
instance, were largely unknown before the 1930s, and were not widely used by
social scientists until the 1950s. The introduction of social surveys in Africa
is much more recent.

Yet even in this short time there have been significant advances in
research methods, a point easily demonstrated by comparing government reports
or academic journals in the 1930s with those published today. If we extra-
polate the rate of advance for another few decades, the collection of quantifi-
able social data will be on a much firmer footing in the near future. This
does not mean that the problems of measurement will be solved, but it does
suggest that the procedures for avoiding self-deception will become increasingly
sound, and quickly so.

Along with the substantial improvements in methods over the past
few decades has come an enormous proliferation of social scientists and social
science writings. This means a great number of people, most of them sensitive
to the complexities of measurement, are working to further improve our research
methods.

Of course neither the past history nor the anticipated advances of
the social sciences necessarily prove their worth. After all, modern weaponry
has become increasingly sophisticated over the past few decades and has
proliferated to nearly every nation, but most of us would prefer a world with-
out this weaponry.

The worth of social science, either as a guide to action or as
source of edification, will in the end be measured by criteria external to
its own enterprise, that is, measured in terms of its creative contribution
in the affairs of men. To make this judgment one must seek out examples
of social science to see if studies have contributed to a better social order.
Our contribution to this task is a very modest one. Here we make the assump-
tion that to evaluate, let alone to use, social science it is necessary to
understand its basic workings. Part, though not all, of these workings **is**
implied by the concept 'research methodology' as discussed on the first page
of the chapter. The present task is to make 'research methodology' under-
standable.

The Purpose Of This Text

This text has been written with two different East African audiences
in mind: the beginning student in the social sciences, and the consumer of
social science writing, especially the civil servant whose job requires a
working familiarity with social science. Neither of these groups practices
social science, but neither can be effective without understanding a few
things about how social science is practiced.

This text is not a 'how to' manual. The emphasis is on conceptual
understanding not procedural knowledge. For instance, we take up the idea of
sampling but we do not go through the computational steps which a sampling
statistician applies as he goes about his task. Instead we review what
sampling statistics are supposed to accomplish. The same is true for the
descriptive and correlational statistics introduced. Less emphasis is placed
on computational knowledge than on conceptual understanding. There are times
when such understanding depends on seeing the properties of certain statistics,
and thus involves computation. But throughout the stress is not in making the
reader a trained social scientist or a practicing statistician so much as
helping him appreciate what the trained social scientist or statistician does.

The student of social science and even more so the consumer of
social science reports in the civil service is forever being presented with
sentences such as the following:

"We took a multi-stage random sample of the farmers in Meru District."

"The correlation between maize yield and inches of rainfall is .63."

"About 35 per cent of the variance in yield is explained by how much rain falls in the area."

"We are confident at the .01 level that the different levels of yield reported are not due to chance."

The typical consumer of research does not have time to learn how to draw a multi-stage sample, or how to compute a correlation coefficient, or how to calculate variance explained, or how to determine confidence levels. But he does want to know what these things mean. He would like to know what data patterns are being summarised by such terms.

We hope that the text will make clear what sentences such as these mean. The material is not always easy. Some of it should be read more than once. And it is preferable to read the text as part of a course where there is an instructor to amplify and further explain concepts and procedures only imperfectly described here. That is, just as this is not a 'how to' manual it is not a 'self-teaching' book. Either of those purposes could only be served with a different sort of book.

Finally we stress that the text is selective. For instance, greater emphasis is given to survey data than to other and equally acceptable methods of data collection. Just as the text is selective, it is limited. We review some of the guidelines in survey questionnaire construction, but not all of the guidelines. Simple correlation is described, but not multiple correlation. There are many textbooks which go beyond the introductory materials covered here, some of which are cited in **bibliographic** notes. The reader who wishes extended discussion or who needs more 'how to do it' instruction is urged to turn to those texts.

CHAPTER 2

WHAT IS A SAMPLE?

All social science rests on sampling decisions, though not all
practitioners recognise this fact.  Persons involved in observation or informal
interviewing often deny that they are in any way 'sampling theorists'.  In
fact, they have made a decision against a systematic sample, but they still
have initiated their research with some rough idea of how many of what kinds
of persons they hope to observe or talk with.

Indeed, sampling is unavoidable in any kind of scientific observation.
The astronomer does not study every single **star, but studies a selected sample**
of stars and then generalises about an entire galaxy.  The historian can seldom
read every document or talk to every old man, but reads selected documents
and talks with selected informants and then attempts to describe an entire
historical period.  The anthropologist wants to do an ethnography of a typical
village so that he can tell us about the customs and practices of an entire
clan or tribe.  The archeologist devises methods for making sure that a
particular potsherd is representative of other potsherds found in that site.
Even the political scientist who relies on a network of informants to provide
insight into the politics of Parliament finds ways of checking the accuracy of
the information provided, and one frequent method is to sample other sources
as **a cross-check.**

Sampling is basic to science because the scientist wants to comment
on broader patterns or more extensive social behaviour than he himself can
ever hope to observe directly.  And although in this chapter we talk primarily
about the sample survey because this is the most common type of sample in social
science, the basic principles we review are applicable to any kind of sampling.

What is a sample?  It is a small part of a large population which is
thought to be representative of the larger population.  Any statements made
about the sample should also be true of the population.  The sample survey,
then, is a relatively cheap and easy way to collect information about a large
group of persons.  If you can learn something about a large group by studying
only a few of its members, then you have saved time and money.  What is

necessary is to be able to <u>generalise</u> from the sample to the population.
Stated differently, what is true of the sample should also be true of the
population. If the unemployment rate of a sample of primary school leavers
is 60 per cent, then the unemployment rate for the population from which this
sample is drawn should also be 60 per cent. Of course if the sample is
incorrectly drawn, the generalisation will be faulty. What is true of the
sample <u>will</u> <u>not</u> be true of the population. In this case we say that the
sample is not representative, and should not be generalised to the population.

What characteristics should a good sample have? Imagine a company
that manufactures jars of jam and a government inspector who wants to find
out whether this jam meets government standards of purity. To determine if
the jam is pure, he must open a jar and test it. But he cannot open all the
thousands of jars that are produced by the company, for that would be totally
impractical from his standpoint and would, moreover, leave the company with
no unopened jars to sell.

Thus, he decides to take a sample of the jars and check them just
as a researcher takes a sample of citizens. What kind of a sample does he
aim for?

(1) <u>An adequate number</u>: The government inspector will probably not
be satisfied by looking at one or two jars of jam. Any particular jar may
be unrepresentative of all the others. He will want a sufficient number so
that he is likely to find contamination if it exists.

(2) <u>A wide geographical spread</u>: If the company produces jam at a
large number of factories in the country, the government inspector should not
be satisfied to go to that factory nearest his own office, even though that
saves him time and effort. The company may be meeting high standards in that
particular plant, but that factory may be different from factories elsewhere
in the country.

(3) <u>A wide range of types</u>: If the company makes all kinds of jam,
the inspector would do well to sample the various kinds and not to limit his
inspection to strawberry jam. Again, the reason is simple: what you find out
about the strawberries may not tell you about the raspberries.

(4) <u>A sample selected only by the investigator</u>: The clever inspector
will be careful to avoid systematic bias in the particular jars he looks at.
He will not sit in his office and allow company officials to bring him a few
examples of each of their kinds of product. Rather he will go himself and
select the jars to be inspected from the regular production of the company.
If he had used the former procedure, there would be a great possibility of
bias since the company officials would bring him only those jars they were
sure were not contaminated.

We can now turn to the sample survey to see how the researcher deals
with these issues:

(1) <u>An adequate number</u>: A good survey researcher will not be satisfied
to talk to one, two or a few dozen citizens. Even if they are selected at random,
they will be too few to allow generalisation to a much larger population. There
is too much chance that one will accidentally find people who differ substantially
from the rest of the public. Public opinion polls **in** the U.S., for instance,
interview about 1,500 respondents. The number is not accidental. **Stati**sticians
can show that with that many interviews, properly selected from a cross-section
of the population, one can estimate within a few per cent the attitudes or voting
behaviour of the entire population.

(2) <u>A geographical spread</u>: It is important that these interviews
come from a wide range of localities, not simply from the researcher's area.
For this purpose, survey organisations maintain staffs of interviewers through-
out the country, and samples are drawn in such a way as to cover a wide range
of geographical areas.

(3) <u>A wide range of types</u>: The most important thing about a sample
is that it be designed to allow all types of citizens an equal chance to fall into
the sample. All sorts of biases can be built into a sample if the criteria used
to select people systematically eliminate one type of person or systematically
lead to the over-representation of another.

A classic example of this phenomenon is found in one of the earliest
attempts to predict a U.S. presidential election on the basis of a sample. In
the 1936 election, a magazine called the <u>Literary Digest</u> **conducted a sample survey**
by telephone on the presidential race between 'Alf' Landon and Franklin D.Roosevelt.

The results predicted a landslide victory for Landon; in fact, the opposite
happened. The **cause of so enormous a miscalculation** was a simple sampling
error. People who had telephones--particularly in the 1930s--came dis-
proportionately from the wealthiest segment of society; these people were more
likely to vote for the Republican Landon. If the Literary Digest had sampled
everybody equally--those with and those without telephones--it might have
predicted the election more precisely.

The mistake made by the Literary Digest is not likely to be repeated
again, for the research community has become much more sophisticated about
the requirements of drawing a suitable sample. The problems more likely to
be encountered now are given by difficult research conditions. It is not
always easy to draw a sample, especially in societies where reliable records
are not kept. And even when records are kept, it may be more difficult to
locate persons than the researcher initially expects.

(4) A sample selected only by the researcher: A further important
aspect of a good sample is that the researcher chooses on the basis of his
statistical criteria who it is that he wants to **interview.** He does not wait
for people to come forward and volunteer to be interviewed. The point is
crucial, and it is this very point that most clearly distinguishes the sample
survey from less demanding methods of studying social behaviour. The survey
attempts to eliminate two kinds of errors. First is the error which occurs
when the respondents select themselves. The government official who estimates
public response to a new programme on the basis of people who volunteer their
opinions gets a very distorted picture. People who volunteer their opinions
usually feel very strongly. Thus the official may be hearing from a small
group of very disgruntled citizens when in fact the large majority are generally
satisfied with the programme. Or he may be hearing from a tiny group of people
who like the programme when in fact the large majority is either indifferent
to it or feels somewhat negative. In either case, his picture of 'public opinion'
is far from accurate.

A related error occurs because people readily generalise from their
own experiences, and do not often realise that their own experiences can be very
limited. One reason for this is that most of us spend our time with persons who

think more or less as we do. Our friends and work associates tend to have
similar social backgrounds and tend to share experiences. Thus, university
students respond in shocked surprise when they are told that large numbers
of citizens view university students as parasites making no contribution to
the nation though draining off much of its valuable resources. After all,
reasons the student, every one he talks with understands that students are
working very hard at their studies so that they will be able to contribute
to nation-building. But of course students talk mostly with other students
or with university faculty and staff. And this set of acquaintances only
reaffirms how important university students are to national development.

The sample survey forces us to consider the viewpoints of a much
wider group than just our own acquaintances. Indeed, when properly conducted,
the sample survey gives every viewpoint in the population an equal chance of
being voiced. Herein lies one of the great strengths of the sample survey.

Simple Random Sample

We have seen that the purpose of a sample is to allow us to make
statements about the population when it is too expensive or unpractical to
collect information from the entire population. The principle of making
statements about the population is called generalising to the population.

We can generalise about the population from results of the sample
if and only if certain conditions are met. The most basic condition is simple:
The probability of any given person falling into the sample should be known.

The easiest way to meet this condition is for everyone's chance to
be equal. If a national lottery sells one million tickets, each ticket purchaser
(if he buys only one ticket) has one in a million chances of being the big
winner. If there is not one winner but 100 winners, then each ticket holder
has $100 \div 1,000,000$, or .0001 probability of falling into the sample of winners.

A simple random sample is one in which everyone has an equal chance
of being sampled, a chance equal to the size of the sample divided by the size
of the population. If a sample is truly random the chance that something will

be observed in the sample is proportional to the frequency with which that something occurs in the population as a whole. Here is an example where the investigators thought they were protecting this basic condition of sampling, but in fact were violating it. As you read through the example see if you can see the flaw before it is pointed out.

This is a study of rural to urban migration. A sample of households has been drawn, and the head of the household is asked to:

1. List every person who has lived in this household for the past five years. For anyone listed who is not presently in the household, tell where they moved to and what work they are doing.

2. For everyone listed, give age and sex.

One research goal is to describe the age and sex composition of persons who leave the household to take up employment elsewhere. We might expect to see in the research report a statement such as: "The average age of persons who migrate from the village to urban centres is 10 years younger that the average age of the population as a whole."

Could you trust this statement? No, because based on this sample you do not know the average age of the population. Why not? Because the basic principle of sampling was violated. Persons who migrate have a greater chance of falling into the sample than persons who do not. Look back at the actual questions asked and you will see why. Anyone who has lived in the household for the past five years is included in the information. A person who has moved within the past five years has twice as many chances of being counted as a person who does not move. If a person moves twice, he has three times as many chances of being counted. Consider the case of Tumwa, who in 1970 left his father's home to work in the near-by town where he lived with his employer. After three years he moved back to his village where he married and started farming. If all three of the households fell into the sample, Tumwa would be counted three times. First his father would report him because he lived in that household within the past five years; then his ex-employer would count him for the same reason; then he would count himself when the investigator came to his household.

Of course it is not likely that all three households would fall into the sample, but this does not change the basic point. The person who moves has a higher probability of being counted than does the person who does not move, and thus the principle of random sampling is violated.

The Sampling Frame

The theory of random sampling is deceptively simple. All you need to do is establish a list of every person in the population from which you wish to draw a sample, and then use some random choice method to select the appropriate number of names to be included in the sample.

Practice is not so simple as theory in this case. Reliable lists, more technically knows as the 'sampling frame', are not easy to come by. This is true of all countries, but especially of countries such as those of East Africa where public record keeping is not a well established practice.

Certainly no nation keeps a list of its **entire** population. And if such a list were compiled it would immediately be out of date as citizens die and new citizens are born. How to get a list? **Sometimes** social scientists use such lists as are available: a telephone book, a post office registry, a list of taxpayers. The dangers are obvious. Such lists only partially cover the population. And the part of the population covered is not necessarily representative of the entire population. The telephone directory includes only households with telephones; the poorer the country, the fewer the telephones, and the greater the bias if the directory is used as a sampling frame. A post office registry is similarly limited, as is a list of taxpayers.

A researcher in Tanzania comments on the difficulties of using such lists as are available:

> I considered the lists to which I might have access -- tax rolls, church or mosque membership lists, school records **showing** parents of students, and so on -- but found each of them to be too limited. A list of tax-payers, for example, does not include those individuals who manage to defy the government's attempts to tax the head of every household, nor does it include those people who are not expected to pay taxes -- for example, elders, women, and children. **My** efforts turned up no list, or combination of lists, which did not systematically exclude at least some of the kinds of people I wanted to interview. I gave up the idea of drawing a sample from any extant enumeration. (20, p.116)

Of course sometimes the target population is not everyone in the country, but a particular group. And then it is often possible to get a list. A sample of junior agricultural officers could be drawn from a list provided

by the Ministry of Agriculture. A sample of businessmen might be drawn from
Chamber of Commerce members (though of course the researcher would only be
getting a sample of the businessmen who belong to the Chamber of Commerce).
A sample of exporters could be drawn from a list of persons who hold export
licenses.

Not all special groups have such lists, however. Consider even
the following groups: Ph.D. holders, prostitutes, shop-owners, newspaper
readers, rich persons, Muslims, any one of which might be worth research
attention, but none of which are conveniently listed somewhere.

Even if a list of the relevant population exists, the researcher
must be cautious in using it to draw his sample. The list itself may be in-
complete, and incomplete in a manner which is misleading. If so, any sample
taken from the list would incorporate the consequent bias. The editors of
a recent collection of essays on the application and limits of survey research
in Africa stress this point. They note that population lists can be compiled
from faulty or incomplete information, and that such lists can be compiled to
promote special interests.

> Census data, for example, are sometimes padded for political
> reasons. Hospital or school records may be distorted to increase
> chances of receiving additional aid from governmental or other
> funding agencies. Voting lists may contain the names of long-
> dead persons. (20, p.97)

The authors of this passage note that such difficulties are not restricted
to Africa, but that distortions may be more difficult to discover in African
societies because so little is known about the characteristics of African
populations. They conclude that the researcher wishing to draw a random
sample from an African population "may find that he has to spend a great deal
of time and effort simply enumerating the members of the population".

Even when the population can be satisfactorily listed and a sample
drawn, the research difficulties are not over. The failure to find people
in the sample, or to get them to cooperate with the study, can destroy the
sample base. Again this is not unique to Africa, but conditions in African
societies can substantially exaggerate the difficulties of obtaining a high
enough response rate to warrant confidence in the sample estimates.

Mugo Gachuhi and his associates report a study of the impact of a
health course taught in the 'model village' of Kirathimo, Kenya. The list
from which they drew their sample was not difficult to obtain, but finding
the respondents became a major problem. They provide an excellent description
of the difficulties, as well as the solution they adopted. The report is worth
quoting at length.

> It was initially intended to draw a $\frac{1}{3}$ sample of all the women
> who had attended the course at Kirathimo village. Thus, a sample of
> 122 cases was drawn from a total population of 366 past participants.
> Systematic random sampling techniques were used in which every 3rd
> participant was drawn from a random start from a list of past partici-
> pants supplied to us by the management of Kirathimo village.
>
> Upon entering the field, however, it became clear to us that
> our sample was destined to suffer an inordinately high rate of attrition.
> Normal attrition rates in East Africa average about 15%. To account for
> this expected attrition rate we had over-sampled by 15%. It became clear
> to us very early that we had severely underestimated the expected attrition
> rate which now promised to be in the region of 66 percent. The sources of
> attrition were as follows.
>
> 1.  Many of the mothers drawn into the sample were unknown in
>     the village from which they were reputed to have come. In
>     some cases, they were no longer remembered by the health
>     visitor who had recruited them.
>
> 2.  Other mothers had emmigrated to parts outside Kiambu
>     District.
>
> 3.  Some had got married or remarried and left to live with
>     their husbands.
>
> 4.  Some of the mothers were inadequately registered at
>     Kirathimo, some of the names used being so common as to
>     make identification well nigh impossible.
>
> Thus, of the original sample of 122, only 43 were definitely
> located and interviewed.
>
> It was, therefore, decided to boost the sample by including as
> many non-sampled participants as could be located during the course of
> interviewing those already included in the sample. A strict record of
> these additional respondents was kept to enable us to separate them
> from, and compare them with the originally drawn sample, thereby enabling
> us to decide whether to treat the two groups separately as being
> heterogeneous or together as a single homogeneous group.
>
> As it turned out, we decided to treat both groups as one single
> homogeneous group. This decision was based on a careful examination
> of the characteristics of each of the two groups to determine the
> degree to which they were different from each other.

An important reservation must nevertheless be recorded. The sample which we finally ended up with is representative more of these ex-participants who could be found and less of those who could not. Unfortunately, we have no way of estimating the degree to which our present sample is or is not representative of those participants who could not be found in as much as our information about the characteristics of those who could not be found is far too scanty. (8)

Gachuhi and his associates here alert us to the practical difficulties which face a researcher hoping to apply random sample techniques. Until the researcher enumerates the population in a manner which insures inclusion of all relevant persons and also insures that the sampled respondents can be located, the procedures of random sampling cannot be applied.

What this implies, in effect, is that the researcher himself must often take responsibility for compiling the basic list from which his sample is to be drawn. Only in this manner can he guarantee that he will be able to find the persons finally to be sampled.

Such a listing operation can be enormously expensive. In fact it would be prohibitively expensive if the study were of the entire population of a society. No research team, however well-funded, can compile a list of the whole population. This is a task even governments are reluctant to undertake, and usually do so only every ten or so years when a census is attempted.

Social science clearly needs some method which allows a sample to be taken but which does not require the prior enumeration of a complete population. The method which best protects the principles of random sampling, but which greatly reduces the costs of a complete listing, is known as multi-stage sampling.

## Multi-Stage Sampling

Multi-stage sampling is just what its name implies, sampling in stages. It is an important modification of simple random sampling, a modification designed to reduce the costs of creating reliable sampling frames. Suppose a researcher wishes to study the health of secondary school students. It would be difficult and certainly expensive to compile a list of every secondary school pupil in the country, but it is not so difficult to get a list of every secondary school in the country. Such can normally be obtained from Ministry of Education officials.

Let us assume that the list includes 700 separate schools, with an average enrollment of 600 students. To take a straight random sample would require listing 600 times 700 students, or 420,000 separate individuals - a time-consuming, expensive task.

The researcher avoids this task by applying the techniques of multi-stage sampling. He first takes a 5 per cent random sample of schools, which gives him a list of 35 schools for his sample. This is the first stage. He visits each school, lists every student enrolled, and then draws a 10 per cent sample of students. This is the second stage. When he has visited all 35 schools he will have listed 21,000 students (instead of 420,000) in order to draw a sample of 2,100 respondents (35 x 600, divided by 10).

Multi-Stage Area Sampling: The most common type of multi-stage sampling when an entire population is being studied involves area sampling as the first stage. Area sampling rests on the simple (if not always correct) assumption that people live somewhere. If areas are sampled, then everyone has an opportunity to fall into the sample. In area sampling the researcher first draws a sample of designated areas, perhaps city blocks or rural locations. Each household in the sampled area is listed, and from that list is drawn the final sample.

This can happen in two or more stages. In Kenya one might first take a sample of districts (first stage), then of locations (second stage), then of sub-locations (third stage), and then of households (fourth stage). If not households but individuals were the final target of the sample, a fifth stage would be involved. But this would require another listing operation. Not only would each household in the sampled sub-locations have to be listed, but each member of the households would be listed (or each member over a specified age if only the adult population were being studied).

The distinction between household samples and population samples is an important one, and is sometimes not noted by researchers. The household sample is often used because listing households within a specific village or on a given city block is very much easier than listing all the people who actually live in that area, and the latter is necessary if a population sample is taken.

Usually in a household survey the respondent sought is the head of the household. But the careless researcher sometimes forgets that a household sample is not a population sample. He asks questions of the head of the household as if the head were typical of the entire population, and reports the results accordingly. For instance, a household sample in which heads of households are the respondents turns up the finding that 70 per cent of the respondents listen to national news on the radio. This finding is reported as if 70 per cent of the population listens to the news. This is misleading, for the head of the household is not typical of the population. The head is more likely to be older, to be male and to be employed than are other members of the population. The only reliable way to get an estimate of radio listenership in the population is to draw a population sample, not try to guess it from a household sample. In short, the population about which the researcher can generalise is determined by the way he which he draws his sample.

Area sampling is widely used because of its lesser costs in comparison with simple random sampling. One reason for the lesser costs has been mentioned, the costs of listing. Another cost reduction is in data collection. It is cheaper to interview 1,500 people living in 100 specified rural areas than to interview the same number of people scattered across the entire countryside.

But area sampling, or any form of multi-stage sampling, has one major disadvantage in comparison with simple random sampling. It takes a larger multi-stage sample than a simple random sample to achieve the same degree of precision in making estimates about the population. (The issue of sample size and sample error is reviewed below).

Stratified Sampling

In random sampling, whether simple or multi-stage, every person in the population has an <u>equal</u> chance and thus of course a known chance of falling into the sample. What is known as stratified sampling represents a departure from the principle of equal chance, though this form of sampling protects the principle of known chance.

In stratified sampling some persons have a greater chance of being selected than other persons. Because these unequal chances, or probabilities, are deliberately selected, the chances are still known. Thus, through statistical

manipulation, usually by assigning different weights to different parts of
the sample, it is still possible to generalise from the sample to the popula-
tion.

Why might a researcher wish to assign unequal chances to be sampled?
The question is best answered through example.  In a study of academic per-
formance in the first year of law school, the researcher wishes to investigate
the relationship between first year exam marks and the type of secondary
school attended by the students.  There are 300 students, 200 of whom attended
high-prestige national secondary schools, 80 of whom attended government-aided
local secondary schools, and 20 of whom attended non-aided secondary schools.
The researcher can afford to study only 60 students, a sampling ratio of 1 in
5.

Taking a simple random sample would produce the following group to
interview:

| | | |
|---|---|---|
| National Secondary School Products | = | 40 |
| Govt. Aided Secondary School Products | = | 13 |
| Non-aided Secondary School Products | = | 4 |

The researcher knows that 4 students in one of his classifications is too few
for sound analysis.  Thus he decides to alter the sampling ratio so as to get
an equal number of students from each type of secondary school.

| | Number in Population | New Sampling Ratio | Number in Sample |
|---|---|---|---|
| National Secondary Schools | 200 | 1/10 | 20 |
| Govt. Aided Secondary Schools | 80 | 1/4 | 20 |
| Non-aided Secondary Schools | 20 | 1/1 | 20 |

The researcher can now compare his three groups, though of course the
sample is no longer representative of the population from which it was drawn.
There can be no generalisations to the population without making adjustments
for the manner in which the 'equal chance' principle was modified.  But because
the 'known chance' principle was not violated, these adjustments are possible.

A stratified sample is one in which the researcher first stratifies
(or clusters, as it is sometimes called) the population according to some pre-
determined criteria, and then samples separately from each stratum.  The purpose

is to insure that enough cases of each stratum fall into the sample to make
analysis possible.  A study of religious practices might over-sample members
of minority religious groups to insure having some of them in the sample.  A
general population study in a country where the population is unevenly distri-
buted might over-sample the sparsely settled areas to insure that its residents
appear in sufficient number to be studied.  A study of birth control might
over-sample the very largest families, which are few in number, to compare
them with the average size families.

## Quota Sampling

The sampling procedures just reviewed -- simple random, multi-stage
and stratified sampling -- are all forms of probability sampling, the technical
meaning of which need not be reviewed here.  It is sufficient to repeat that
in these forms of sampling the chances of any given individual in the population
being selected in the sample are known.

The **final** type of sampling procedure reviewed, quota sampling, is
not a probability sample.  There is no way of estimating the chances of any
given individual being sampled, and indeed no assurance that every type of
individual has some chance of being included.  Thus the possible error in
generalising from a quota sample to a population is very much greater than
when generalising from probability samples.

Given these limitations, the reader might well wonder why quota
sampling is included in our discussion.  The reason is simple.  Quota samples
are widely used; they are especially widely used in Africa where the costs and
difficulties of listing (a basic requirement of probability sampling) can be
prohibitive.

In quota sampling the researcher attempts to select a sample which is
a replica of the population he is studying.  For instance, if the population is
half male and half female, then the sample should be half male and half female.
If the population is 60 per cent literate and 40 per cent illiterate, then the
sample should have the same distribution.  And the criteria of sex and literacy
should be combined, so that the sample has the appropriate proportions of
literate males, literate females, illiterate males and illiterate females.

Say that it is known about the population that it is composed of the following four groups:

| | |
|---|---|
| Literate Males | 40% |
| Literate Females | 20% |
| Illiterate Males | 10% |
| Illiterate Females | 30% |

If sex and literacy are the criteria to be used in taking the quota sample, and if the researcher wants a sample of 1,000 persons; he will seek out 400 literate males, 200 literate females, 100 illiterate males, and 300 illiterate females.

The conditions of his quota sample would be fulfilled, but the sample may not be representative of the population as regards other dimensions. For instance, it may only include middle-aged people (who often are easier to find than the young or the old), and thus be unrepresentative as regards the age structure of the population. Of course age could be added as a criterion of the quota sample, but this would further complicate matters. And at some point the researcher has to stop adding criteria. For instance, there is no way he could sample a quota of innovative and non-innovative farmers, for there is no way to recognise these traits short of carrying out the research itself.

The difficulties and limitations of quota samples are numerous. But often there are no alternatives, and a carefully thought-out quota sample is superior to haphazard interviewing of whoever happens to be available to the researcher. An imaginative use of quota sampling is illustrated in a recent study of fourteen parliamentary constituencies in Kenya, carried out under the direction of John Okumu and Joel Barkan. They describe their sampling procedures as follows:

> We took a quota sample given the impossibility of obtaining or creating lists of the populations of the fourteen constituencies in which we conducted the surveys. To insure a representative sample, we sought to interview 300 citizens in each constituency—the maximum we could afford, and a number which if selected randomly would have resulted in an error rate of 5.5%.

> On the basis of the 1969 census, quotas were set according to age and sex. These quotas were then randomly distributed across each constituency. The virtue of the scheme, of course, was that relatively novice interviewers could quickly locate suitable

respondents, something which would have been extremely difficult
even if a strictly random procedure had been used. Equally
important, we could verify each interview, because they were
clustered in groups of ten, and each plot (about a quarter mile
in diameter) was clearly marked on the map.

Each interviewer, moreover, was given a check-off sheet which
indicated the identification numbers on each of his questionnaires,
the age and sex of the respondent for each questionnaire, and the
plot in which each questionnaire should be used. Each interviewer
was also given a map on which the plots were clearly marked and
numbered. Each was assigned 100 interviews, ten interviews in
each of ten plots.

Another virtue of this system was that because each interviewer
had to move around on foot, the distances they had to travel were
within feasible limits. Again, this would not have been possible
had we attempted a random selection process.

Surveys in rural areas of developing countries present a whole set
of logistic problems not encountered in developed countries.
Albeit not a perfect substitute for random sampling, we think the
above method is about as good as one can devise. (1)

## How Large Should A Sample Be?

To estimate a population characteristic of the United States with its
125 million adults requires a sample of approximately 1,500 persons, and to
estimate a characteristic of the Kenyan adult population of 8 million at the
same level of accuracy also requires a sample of approximately 1,500 persons.

This fact should quickly dissuade the reader of the common misunder-
standing that the larger the population the larger the sample needed. Other
factors have much more bearing on sample size than does the size of the population
itself. In fact, whenever the sample is less than 5 per cent of the population
(which is true of nearly any large survey study) sample size need not be at all
related to population size. This is why a sample of 1,500 works just as well
in the United States with its 125 million adults as it does in Kenya with its 8
million adults.

If it is not the case that the larger the population the larger the
sample needed, what are the criteria of sample size? One often-cited criterion
is heterogeneity. The more heterogeneous the population, the larger the sample
needed. If a population is identical on the trait being estimated, then a
sample of one is adequate. For example, the number of blood corpuscles in a

blood sample taken by a doctor is minute compared to the total number in the body. But the sample can be counted on to be representative because the composition of one blood sample is like that of **any** other from the same patient.

Yet the rule about homogeneity is not of much practical value. For one thing, we usually do not know whether human populations are homogeneous or not. Is a population of students homogeneous or heterogeneous with regard to intellect? What about farmers with regard to use of particular seeds or fertilisers? What about civil servants with respect to efficiency? The reason for taking a sample of students or farmers or civil servants is that we want to learn how the given trait is distributed.

There is a second reason why homogeneity is not a practical guide to sample size. In any study we want to measure the population on more than one characteristic. And whereas on one characteristic the population may be fairly homogeneous (crops grown on a particular soil), on other characteristics it may be very heterogeneous (how much fertiliser used, how much the farmers are in debt, size of families, etc.).

So if heterogeneity/homogeneity of population is not a practical guide, what is? The fact is that there is no simple guide. The size of the sample required varies with many things - with the costs of the study, with the homogeneity of the population, with the number of traits to be measured, with the kind of sample drawn (simple random, multi-stage, stratified) and with the size of sample error the investigator is willing to tolerate.

The notion of sample error introduces a new idea into our discussion. It is an important idea, one meriting extended review. Unfortunately, however, a thorough understanding of sample error requires some background in statistics. Some of the relevant statistical concepts are reviewed in later chapters, and the reader may wish to re-read the following material after those concepts have been learned.

For the present a very general introduction to sampling statistics will be sufficient. As stated earlier, a sample is taken in order to make an estimate about the characteristics of a population. A good sample is one which accurately estimates the distribution of the characteristic in the population; a poor sample inaccurately estimates the distribution.

Sampling statistics can accomplish two things for us. First, they can determine the range within which our sample accurately estimates the population characteristics. Second, they can tell us how much confidence to place in the estimate. To understand both of these ideas is to understand much that is important about sampling.

Sampling Error: We draw a sample in order to estimate a characteristic of the population. Can we trust our estimate? This is the nagging question which faces any scientist who uses sampling techniques. Say that a sample of farmers is questioned and it is learned that 30 per cent of them are using a new, highly reliable fertiliser. But are 30 per cent of all the farmers in the region using the fertiliser? Perhaps our sample is 5 per cent off, and the actual proportion in the population is only 25 per cent or as much as 35 per cent. If this were the case, we would say that there is a 5 per cent sampling error. The population characteristic is within plus or minus 5 per cent of the sample estimate.

The sampling error may have been larger or smaller. For instance, if the sampling error were 10 per cent, the researcher would have to conclude that somewhere between 20 and 40 per cent of the farmers had adopted the new fertiliser, but if the error were only 2 per cent he would conclude that no fewer than 28 per cent would have adopted the fertiliser.

Let us consider how one might make a decision regarding tolerable sampling error. If the study of fertiliser adoption were an exploratory study in which the researcher simply wanted a rough guess about the rate of adoption in different regions, then a sampling error as high as 5 or 10 per cent might be acceptable. If, instead, this study were to be used to plan a massive and expensive government programme of fertiliser distribution, then a good deal more precision might be called for.

Let us assume that social research has demonstrated that when one-third of the farmers in a region have adopted a new farming technique the remainder of the farmers will adopt it in the near future. That is, one-third is a threshhold point at which a new technology is rapidly diffused. But when fewer than one-third have adopted, then the other farmers resist the new technology. They feel it has not sufficiently proven itself.

If the researcher had designed his study with a 5 per cent sampling error tolerated then he is not in a position to provide good advice to government. A region with a 30 per cent adoption rate according to his sample might actually have a 35 percent adoption rate, and thus not need the government programme. Or it might have an adoption rate as low as 25 per cent, and thus clearly stand in need of the government programme. Had the research been designed to estimate more precisely, say within 1 or 2 per cent, then the researcher could have reported to the government planning unit that the region with 30 per cent adoption according to the sample almost certainly needed the government programme.

Sampling error, then, is the range within which the population characteristic probably falls. A large sample error means that the estimate of the population characteristic is just a rough guess, and that it can be off quite a bit. A small sample error means that the estimate is much more precise, perhaps within 1 per cent of the population characteristic.

The reader might at this stage be asking why anyone would ever design a study which tolerated a large sampling error. The answer is simple: study costs. We will see below that the size of the sample has to get very large in order to reduce sampling error to a low percentage, and large samples are much more expensive than small samples.

Degree of Confidence: When we use a sample we do not know for certain what the exact proportion of the population is which has a certain characteristic, but we can estimate the range within which the population characteristic will fall. This was the first idea introduced, and it was called the sampling error. Now we confront a second issue. It so happens that we are not 100 per cent certain that the population characteristic does fall within the range established by the sampling error. We are probably correct, but not absolutely so.

The 'probably' can be translated into what sampling theorists call the 'confidence interval'. There are statistical procedures for helping us decide how much confidence to place in our estimate. Let us assume that in the above example we designed the study so that it would have a 2 per cent sampling error. Thus we estimate fertiliser adoption to be between 28 and 32 per cent. How sure are we that the adoption rate really is within this range?

We can vary the degree of confidence we want by varying the sample size. If we took a really large sample we could be certain 999 times out of 1,000 that the adoption rate fell within plus or minus 2 per cent of the sample estimate. But such a sample would be extremely expensive. By convention, most social scientists use either the 95 per cent or 99 per cent confidence level. If the former, the researcher would conclude that he is 95 per cent certain that the true population characteristic falls within the range established by his sample error. If he chooses the more demanding confidence level of 99 per cent, then he is even more certain that his estimate is accurate. The more demanding the confidence level desired, the larger the sample needed.

The idea of confidence interval (or confidence level as it is sometimes called) is often confusing to the beginning student of social science methodology. We can return to our example of fertiliser adoption to see how the confidence interval works.

The investigator wants to be fairly precise so he designs his study to have a sampling error no greater than 2 per cent. He conducts the research and reports to government that somewhere between 28 and 32 per cent of the farmers have adopted the fertiliser. He reports that he cannot be more exact, because the size of the sample he took allows a margin for being off by 2 per cent, either higher or lower, than his sample estimate. The official who reads his report, and must make the decision about whether to include the region studied in the fertiliser distribution programme, asks the researcher, **"But** how sure are you that the sample estimate is reliable? It is going to cost us a lot of money to put our programme into that region, and we don't want to do it unless we are certain that fewer than one-third of the farmers have adopted the fertiliser."

The researcher replies, "Well, I am 95 per cent certain. I cannot be more certain than that because of the way in which the sample is designed."

The official then says, "What does it mean to say that you are 95 per cent certain?"

For a person who understands statistical reasoning, this **is** a simple question to answer. But for the untrained person, it is a very difficult one. Let us see how the researcher answers.

"I took a sample to try to learn how many farmers in the region were using the new fertiliser.  Thirty per cent of the farmers in the sample replied that they were using it.  Knowing what I know about the sample design, I am 95 per cent confident that this is an accurate estimate of the population characteristic (within the sampling error allowed of plus or minus two per cent).

"This 95 per cent confidence level means that if I had taken <u>100 samples</u> instead of only one sample, then in 95 of these samples we would have found between 28 and 32 per cent of the farmers to be fertiliser adopters.  However, in 5 of these samples we would have found the number of adopters to be either less than 28 per cent or more than 32 per cent.

"Of course, I did not take 100 samples.  I took only one.  And I cannot know for certain whether the one I took is part of the 95 or part of the remaining 5.  If the sample actually drawn is part of the 5, then I'm wrong in telling you that the population adoption rate is within the specified range.  If the sample actually drawn is part of the 95, then I'm right.  But I have no way of knowing whether my single sample is part of the five or part of the 95 hypothetical samples.  This is why I have to say that I'm only 95 per cent confident of my finding."

Table 1 shows the sample size needed to achieve a given sampling error at both the 95 per cent and 99 per cent confidence level.  This table represents sample size needed for simple random sampling; it is <u>not</u> applicable to different types of sampling.  A multi-stage sample, for instance, where one first samples geographic areas and then takes household samples within a selected number of areas would require larger samples in order to achieve the same level of confidence and precision.  For example, if one wanted a 2 per cent sampling error and wanted to be confident at the 95 per cent level in a two-stage sample, the sample size would have to be closer to 4,000 than the 2,401 listed in the table.

The table is intended only to illustrate the relationship between error margins, confidence levels and sample size.  It should <u>not</u> replace a statistical understanding of sampling theory if the reader actually wishes to draw a sample that meets stipulated criteria.

Table 1. Sample Size Which Produces Different Levels of Precision and
of Confidence: For Random Samples.

|  | Sample Size Necessary If You Want To Be | |
| --- | --- | --- |
| Tolerable Sampling Error is Plus or Minus | 95 percent certain of your sample estimate | 99 percent certain of your sample estimate |
| 1% | 9,604 | 16,587 |
| 2% | 2,401 | 4,147 |
| 3% | 1,067 | 1,843 |
| 4% | 600 | 1,037 |
| 5% | 384 | 663 |
| 6% | 267 | 461 |
| 7% | 196 | 339 |

If we return to our example we can see how the sample error and the
confidence level interact. The researcher reported 30 per cent of the farmers
to be using the new fertiliser. Assume that his sample was 700 farmers. He
could report one of the following to government:

> The rate of adoption of the fertiliser among the farmers of the
> region studied is between 26 and 34 per cent. And I am 95 per
> cent confident that I am correct in this estimate. Though I do
> warn you that there is a 5 per cent chance that I am wrong.

> or

> The rate of adoption of the fertiliser among the farmers of the
> region studied is between 25 and 35 per cent. And I am 99 per
> cent confident that I am correct in this estimate. Though I do
> warn you that there is a one per cent chance that I am wrong.

Either statement is an accurate reflection of his finding given the size of
the sample he used (if it were a simple random sample).

As you can see, it takes a very large sample to reduce the sampling
error to less than three per cent. And the sample has to be larger if the
confidence level is set at 99 rather than 95 per cent. For instance, if the
government planning unit had wanted a sample estimate correct to within one
percentage point and one in which they could be 99 per cent confident of the
results, the sample would have had to have been 16,587 or nearly 28 times as
large as the sample we have hypothesised (and probably more than 28 times as
costly).

CHAPTER  3

THE  SAMPLE SURVEY

The first chapter distinguished reactive from non-reactive research. An example of non-reactive research is research which uses public documents as a source of data. In such research there is no way to create new information; one takes the information he can get. An example of reactive research is the sample survey. In this type of research one deliberately provokes a reaction or response from the person studied. This enables the researcher to create new information. The new information is the answer a respondent gives to survey questions.

The ability to create new information is the enormous strength of a sample survey. It distinguishes the sample survey from many other research designs. The premise  of the survey technique is a simple one: if you want to know something about people, why not ask them? We learn about the finances of the household by asking a member of the household; we learn what a woman thinks about birth control by asking her. Of course there are other methods of learning about household finances or birth control attitudes, but alternative methods to the sample survey seldom provide direct answers relatively cheaply from a large sample of households.

Yet if the basic premise of the social survey is simple, the operations involved in using a sample survey research design are not. This we have already begun to realise in the discussion of sampling theory. Now we will see that the operations involved in writing a questionnaire and administering it to the sample are equally complex.

In fact, drawing a sample is perhaps less complex, or at least less ambiguous, than collecting the desired data from the respondents. Sampling is a science. There are specific rules to follow. If these rules are observed, the sample will have identifiable characteristics (as we saw, there is a direct relationship between size of sample and size of sampling error). The rules of sampling can be learned and applied by anyone with an elementary understanding of mathematics.

Designing the questionnaire to use in a survey is <u>not</u> a science.
There are general guidelines to follow, but no specific set of rules.  The
researcher chooses from among dozens of different questionnaire formats and
several different interviewing techniques, and is never certain what the
result will be of choosing one approach rather than another.

Why is drawing a sample scientific but designing a survey question-
naire not?  Because in sampling we deal with the world of numbers.  Numbers
have definite properties and when used correctly lead to predictable results.
The construction of a questionnaire and planning of an interview takes
research out of the world of numbers and into the world of human interaction.
Humans, especially when reacting to a stranger armed with a lot of questions,
are not predictable at all.  While one respondent may be flattered to have
his views solicited, the next respondent may resent the time an interview
takes and a third respondent may be suspicious of the motives of the researcher.
The first respondent gives misleading answers because he wants to please the
interviewer, and thus emphasises what he thinks the interviewer wants to hear.
The next respondent gives misleading answers because he wants to rid himself
of the interviewer, and thus answers in such a way as to shorten the interview.
The final respondent gives misleading answers because he fears the interviewer,
and thus hides or distorts the true facts.

The problem is that there is no way to predict when a respondent will
be flattered, when he will be resentful and when he will be suspicious.  A
person designing a questionnaire must anticipate all three kinds of respondents,
and many more besides.  Only a well-constructed questionnaire can compensate
for the possibility of misleading information.

But to return to the earlier point, designing a well-constructed
questionnaire does not follow from the application of some exact rules of
science (though drawing a well-constructed sample does follow from the correct
application of known rules of science).  Even the seemingly most straight-forward
question asked of cooperative  respondents cannot escape problems.

Let us take an illustrative question aksed on a standard household
survey in a rural area.  Let us further assume five respondents, all of whom
want to be cooperative and all of whom believe they are correctly answering
the question.

"What crops are you growing here on your farm?"

This is a coffee region, and the research team wants to learn whether farmers are growing coffee as a cash crop and whether they are also growing vegetables for home consumption. All five farmers produce both coffee and vegetables. But here are the answers given to the question.

Interviewer: "What crops are you growing here on your farm?"

| Farmer | Answer | How the Farmer Interpreted the Question |
|---|---|---|
| 1 | "Coffee and vegetables" | This farmer has correctly interpreted the question. |
| 2 | "I am not growing any crops." | This farmer takes the word "you" literally; and because his wife and daughters take care of the shamba while he takes care of the livestock, he answers as truthfully as he can. |
| 3 | "A few vegetables" | This farmer takes another part of the question too literally, the part which says "here on your farm." Although he grows coffee, he does so on a rented plot several kilometers away and he does not really think of that plot as being part of his farm. |
| 4 | "We have an acre of coffee." | Although this farmer does grow vegetables, he does not mention them because he normally thinks of the word "crops" as meaning only cash crops. Because he does not market his vegetables, he fails to tell the interviewer about them. |
| 5 | "Just coffee" | The interview takes place after this farmer has harvested his vegetables, but before the vegetable planting season. He thinks to himself, "Well, I am not actually growing any vegetables right now, so it would not be correct to mention anything other than the coffee plants". |

We admit that this is an exaggerated example, but it is less atypical than one might believe. All five of the answers are perfectly reasonable, given the interpretation of the question. And if the interviewer simply mechanically wrote down what he was told, the results reported by the researcher would be greatly different from the true conditions. The table below contrasts the reported data with the true data.

| | Percent of Farmers Who | | | |
|---|---|---|---|---|
| | Grow Vegetables & Coffee | Grow Coffee Only | Grow Vegetables Only | Grow Neither |
| Reported Data: | 20% | 40% | 20% | 20% |
| True Data: | 100% | 0% | 0% | 0% |

People reading the reported data would draw very misleading conclusions. A nutritionist, for instance, might conclude that the children in this area had poor diets because so few farm families grow vegetables for home consumption. An agricultural economist might conclude that the marketing system must be very in-effective **because** so few farmers are growing the main cash crop for the region. A social scientist who specialises in development administration might conclude that the extension officer must not be doing his task well, or more farmers would be growing both coffee and vegetables.

A very different kind of conclusion would be drawn by a person who knows the region well. Because this local expert knows that most farm households in this region grow both coffee and vegetables, he concludes that survey research is use-less.

Actually, the local expert is wrong in coming to this conclusion, just as the nutritionist, the agricultural economist and the development administration scholar are all wrong in their conclusions.

Survey research is not useless, though the results of this particular survey are. Survey research is a tool, just as a typewriter is a tool. And like a typewriter, it can be used effectively or ineffectively. An untrained typist makes so many mistakes that one is just as well off with handwritten material. An untrained survey researcher makes so many mistakes with his tool that one is just as well off with casually collected data. But the trained typist correctly using the typewriter can produce a large amount of easily readable material in a relatively short time. Likewise, a trained survey researcher correctly using the survey tools available can produce a large amount of reliable information in a relatively short time. The problem is not the tool; the problem is the correct use of it.

The Survey Instrument: Three Basic Choices

Before actually designing a survey questionnaire the researcher has to make three basic choices. First, he decides whether the study will use a self-administered questionnaire or a direct interview. Second, he decides whether the questionnaire or interview will be structured or unstructured. Third, he decides whether to use open or closed questions. These terms become clear as we proceed. We should point out that the choices are not necessarily either/or;

one can mix open and closed questions in the same interview, use a partially structured and a partially unstructured questionnaire and even design a study that includes interviewing and self-administered questionnaires. However one might mix these different approaches, it will still be important to understand the differences implied by the three basic choices.

## The Self-Administered Questionnaire and the Interview

There are two very general ways in which to conduct a survey. The research can prepare a questionnaire which the respondent himself reads and answers. The self-administered questionnaire is analogous to an examination in which the pupil reads the questions and then either chooses the correct answer from among those presented by the examiner or drafts an answer in his own words. In the interview, the researcher, or someone designated by the researcher, asks the question to the respondent in a face-to-face setting, and then the interviewer records the answers. Here the analogy would be a job interview.

## Advantages and Disadvantages of the self-administered Questionnaire: Most self-administered questionnaires are either mailed to the respondents or are distributed directly to them. An example of the former would be if the researcher mailed a questionnaire to every member of Parliament, and requested them to fill in the answers and return it to him (normally he would supply a stamped, self-addressed envelope). An example of the direct distribution technique is common in studies of school children. The researcher goes to the school, distributes the questionnaires (often in a classroom), and then waits to collect them when they have been completed.

The major advantage of a self-administered questionnaire, whether mailed or directly distributed, is coverage. Information from a large sample can be collected relatively inexpensively. For instance, one can get information from a class of 200 students in an hour or two. To get the same information from an interview would take two or three hundred hours.

Another advantage is the anonymity that can be provided. In self-administered questionnaires one seldom asks the respondent to put his name on the questionnaire. Thus he is more certain that no one can connect his answers with him personally. This anonymity often produces more candid answers than is possible in an interview.

But the disadvantages of the self-administered questionnaire probably
outweigh its advantages in an East African country. The simple reason is literacy.
Only a literate population can respond to a self-administered questionnaire. And
thus such a technique is automatically ruled out for national samples in countries
with high illiteracy rates, though of course the self-administered questionnaire
can be used for special groups if all the members are literate. Secondary
school leavers, for instance, have often been the target of surveys using
self-administered questionnaires. It would be extremely expensive to trace
a sample of secondary school leavers all over the country in order to interview
them; thus the mailed questionnaire is used when a large sample is desired
but the study cannot pay the costs of many interviewers.

The use of mailed questionnaires to special target groups introduces
another major disadvantage however. This is the problem of response rate.
Response rate refers to the number of people in the sample who actually answer
the questions. It is a problem in interview studies as well, but it is a much
greater problem in studies which use self-administered questionnaires.

For various reasons people do not return mailed questionnaires.
Sometimes they never receive them - this happens especially in the case of women
who have married and changed their names and is an important problem in school
leaver studies which use mailed questionnaires. Others receive the questionnaires
but either misplace them or throw them away. It is much easier to refuse a
researcher who contacts you through the mail than to refuse a researcher standing
at your front door.

Indeed, even when sampling a literate group whose cooperation one has
taken great pains to secure, the response rate to a mailed questionnaire is
often less than 50 per cent. This destroys any possibility of using the sample
to generalise to the population, for there is no way of knowing if the non-
respondents might differ in some systematic way from the respondents.

A recent study in Uganda demonstrates the difficulty of obtaining
a high response rate from mailed questionnaires. This was a study of school
leavers from the graduating classes of 1954, 1959, 1964 and 1969. The purpose
was to learn whether the more recent school leavers  had greater difficulties
finding employment than had the earlier school leavers. The researcher writes:
"Questionnaires were sent to 2,167 Ugandan secondary school leavers. The

questionnaire dealt mainly with their socio-economic backgrounds and their educational and occupational histories." (6, p.43)  Only 853 of the questionnaires were returned, a response rate of less than 40 per cent.  This is not uncommon in studies which use mailed questionnaires, even when the researcher, as in this example, has the support of authorities and sends several reminders to the sample.

There is a further difficulty with self-administered questionnaires. The researcher loses control over the research setting.  If one mails out a questionnaire and it is returned with all the answers marked, there is really no way of knowing whether the person in the sample actually completed the questionnaire himself.  Busy politicians and businessmen, for example, often hand over questionnaires to their secretaries, telling them to fill them in as best they can.  Related to this possibility is the problem of misunderstood questions.  In an interview one can often sense that a question has not been properly understood and can rephrase it to make it more intelligible.  This flexibility is absent in the self-administered questionnaire.  Take the example cited above.  A quick interviewer might have noted that the question "What crops are you growing here on your farm?" was getting odd replies, and might reformulate the question in the following way: "In the past year what crops have you or members of your household grown on any land that you have the use of?"  There is no way to reformulate a question on a self-administered questionnaire; indeed, there is not even any way of learning whether a given question is being understood by the respondents.

Because of the many difficulties with self-administered questionnaires they are used infrequently in East African countries.  They are used more often in a country such as the United States, where the population is more accustomed to questionnaires and where interviewing costs can run as high as Shs 700 per interview.  Still, even in the United States the problems of low response rate and lack of control over the research setting lead researchers to prefer the interview, when they can afford it.

Advantages and Disadvantages of the Interview:  In the interview the researcher (or an interviewer he has hired) interacts directly with the respondent. Questions are asked and answers are recorded.  The interview provides the researcher with some measure of control over the research setting.  For instance, questions might be modified if it appears that they are being misunderstood (though we shall see that modifying as you go has its own disadvantages).  Control over the research setting is also provided through the judicious use of probes.  A probe is a follow-up question suggested by the respondent's initial answer.  The most common type of probe is simply a request for additional information, as in the following:

QUESTION:  What are the major problems you face in trying to sell your crops?

PROBE:  In addition to the things you have mentioned, can you think of any other problems related to selling your crops?

A probe can also be more directive.  In the following example the
researcher did not want to ask specifically about the subject-matter of the
probe question unless the respondent failed to mention it; if the respondent
did not mention it, then the researcher wanted to direct his attention to the
topic.

>            QUESTION:  What are some of the advantages and disadvantages of
>                       marketing your crops through the local cooperative?
>
>            PROBE;  USE ONLY IF NECESSARY:
>
>                       What about the distance from your farm to the **nearest**
>                       cooperative -- do you find that the cooperative is near
>                       enough to your farm for you to make effective use of
>                       it?

The flexibility provided by such supplementary probes as illustrated above is
an advantage of the interview technique which is denied the researcher using
the self-administered questionnaire.

Another major advantage of the interview is that it is not restricted
to literate populations.  This is one reason why the interview technique
predominates in survey studies conducted in East Africa.

Against the advantages of the interview must be balanced its dis-
advantages.  Foremost among these is cost.  **Inter**viewing is expensive, either
in terms of research time or research money.  If the researcher plans on conducting
all the interviews himself he either is restricted to a small sample or he must
allow a great deal of time for interviewing.  A one-hour interview might
actually take as long as three hours on the average, by the time one adds up
travel time, the time it takes to contact the respondent and solicit his
cooperation and the time it takes to edit a questionnaire after the interview
is complete.  Even when the respondents live fairly close together (which, you
recall, is one reason for using area samples rather than simple random samples)
it is unusual for an interviewer to complete more than two interviews a day.

If the researcher plans to employ interviewers, which is usually the
case in large surveys, then there are the costs of training the interviewers
and of paying their travel and living costs.  This arrangement has costs other
than monetary ones.  The more interviewers there are the more difficult it is
for the researcher to maintain control over the research.  No matter how much
time is spent in training, interviewers will differ in the emphasis they give
to questions and will differ in when and how they probe.  Thus the administration
of the questionnaires is not standardised from one respondent to another, with
the possible result of low quality data and lack of comparability.

A further disadvantage of the interview derives directly from
the interaction between interviewer and respondent. The kind of relation-
ship formed in the interview situation, the way in which questions are asked,
the attitude the interviewer takes toward the research, the carefulness
(or carelessness) with which the interviewer records answers and a host of
related considerations can all introduce distortions in the data.

For example, numerous studies of interviewer bias, as it is techni-
cally called, have demonstrated that women interviewers get systematically
different **answers** to questions on topics such as sexual practices or birth
control measures than do men interviewers. Other traits of the interviewer
which affect response patterns are his or her race and his or her social
status. And of course in East Africa one would expect the tribe of the
interviewer to affect how a respondent answers questions.

The problems of interviewer bias are aggrevated when the research
is being conducted by many different interviewers. And though there are
techniques for reducing the errors introduced in interview studies, these
techniques can by no means eliminate all errors. The human interaction which
occurs **when** an interviewer and respondent spend an hour or two together has
subtle effects on the quality of the data collected. We could not expect it
to be otherwise. But at least the well-trained researcher will know about
these subtle effects and will attempt to guard against the possible resulting
bias in his data.

One final problem of the interview to note: It is difficult to
convince the respondent that his anonymity will be protected. The interviewer
can assure him that his answers will be treated confidentially and will be
used only for legitimate research purposes. But the respondent knows that
the interviewer knows who he is. And if he is suspicious of the interviewer
or of the sponsoring institution, his answers will not be candid, at least
when they touch upon matters he wishes to keep confidential. Even if anonymity
is not an issue, embarrassment might be. This is a problem avoided in self-
administered questionnaires and is one reason why self-administered question-
naires rather than interviews are often used when trying to get information
about matters such as birth control practices. (Though the population from
which one wants birth control information is often not literate enough to
complete self-administered questionnaires.)

No final answer can be given regarding the superiority or **inferiority** of one research technique or the other. The reason that both self-administered questionnaires and interviewing continue to be used by the social science community is that both techniques have their place. The technique chosen will depend on such considerations as size of the sample needed, purposes to be achieved by the data, skills of the researcher and his staff, literacy of the sample, availability of the respondents and, of course, the budget of the study.

The Structured and the Unstructured Questionnaire

A structured questionnaire is one in which every respondent is presented identical or nearly identical questions and the questions are always asked in the same order. The structured questionnaire standardises the stimulus presented to the respondents as much as possible. A structured questionnaire can be used whether one is interviewing or using a self-administered question-naire.

In contrast, the unstructured questionnaire allows the researcher to modify the questions he asks and the order in which he asks them from one respondent to the next. Obviously, the unstructured questionnaire is applicable only to an interview approach. In interviewing with an unstructured question-naire the researcher usually has a definite list of topics he intends to cover, but he will cover these topics in a different **order and using different questions depending on how the interview** situation develops.

The unstructured interview technique is often used by ethnographers as they attempt to develop a coherent and total picture of the customs and beliefs of the people they are studying. Another use is in oral history studies where the goal is to record whatever information an old person has about an historical period or event. Political scientists have made effective use of unstructured interviewing in studies of power in a community or of the role of corruption in politics. It is felt that the direct and more obvious technique of structured interviewing is not always appropriate or wise when investigating such issues.

The general rule is that the less one knows about a topic the more likely one is to rely on unstructured interviewing. It is a method of exploratory research which might then be followed up with more structured interviewing at a later time. The researcher feels he cannot frame intelligent

questions until he has conducted some informal, unstructured interviewing.
In a survey of registered land-owners in the Kisii District of Kenya, one
researcher relied on unstructured interviewing for exactly this reason.  He
writes that in beginning his research "no rigidly structured questionnaire
was used.  My belief that the informal and flexible approach could yield
more fruitful information from the respondent was confirmed in practice.
Much information was volunteered which could not have been anticipated in
the framing of a questionnaire." (23)

However, what the unstructured interview gains in flexibility it
loses in comparability.  As the researcher modifies questions in moving from
one respondent to another it becomes increasingly difficult to compare the
answers of a later respondent with the answers of a previous one.  This is
because the researcher  is changing the stimuli presented to the respondent.
Here is an example of how one loses comparability.

Interview of first farmer:

> "Do you own more than one plot of land?" Answer: "Well, not exactly.
> I own one plot, but then I also share ownership of another plot with
> a brother-in-law."

Interview of second farmer:

> "Do you own more than one plot of land, either on your own or in
> partnership?" Answer: "I own one plot, and I share a plot with my
> uncle.  Then we also have a plot registered in my wife's name."

Interview of third farmer:

> "Do you own more than one plot of land, either on your own or in
> partnership; and is there any land owned by any other members of
> your household?" Answer: "We own three plots of land, one is
> registered in my name and two are registered in my father-in-laws
> name, but I farm all three plots."

The interviewer in this example has changed his question to  take into account
the new information he has obtained in his interviewing.  It is this flexibility
which gradually makes his data more accurately represent the actual facts of
land ownership within each household.

The example also illustrates the price he must pay for this flexibility.
Without going back to interview the first farmer he has no way of knowing whether
that household might not also have included plots which were registered in the
name of members of the household other than the person being interviewed.  And
the researcher cannot constantly be going back to re-interview people every time
he modifies the questions asked.  Rather quickly they will get tired of the fact
that the researcher does not know exactly what he wants to ask them, and they will
stop being cooperative.

In this example, then, there is no way to compare the response of the first farmer with the response of the third farmer, because they were presented such different stimuli. It is not until the question has been standardised that one can compare respondents across an entire sample.

This is the main disadvantage of the unstructured, as compared with the structured, interview. In the latter, a specific set of questions are asked of every respondent. The interviewer does not modify his questions as he goes, although he may use slightly different probes in order to elicit a full response to his questions. The advantage is in the comparability achieved. If every respondent is asked identical questions, then one can compare how different respondents have answered. Indeed, it is difficult to apply quantitative analysis techniques unless the survey instrument is structured, that is, unless every respondent receives approximately the same stimulus.

## Open and Closed Questions

Although there are many different forms of questions which can be included in a survey, most can be classified as either open or closed. An open question is one in which the respondent uses his own words to answer. A closed question in one in which the researcher presents possible answers, and then asks the respondent to select from these answers. The difference can be shown through example.

An Open Question:

What would you say have been the economic results of Uhuru in Kenya?

A Closed Question:

Would you say that Uhuru in Kenya has improved or hindered the economic condition of Kenyan citizens? Choose one answer.

_____ Uhuru has greatly improved the economic condition.

_____ Uhuru has somewhat improved the economic condition.

_____ Uhuru has not made much difference to the economic condition.

_____ Uhuru has somewhat hurt the economic condition.

_____ Uhuru has greatly hurt the economic condition.

Obviously the closed question provides much greater control over the kind of answer one is likely to get. However, the closed question forces the

respondent to choose among alternatives which may not be indicative of his
thinking about the economic results of Uhuru. Say, for instance, that some-
one answered the open question as follows:

> Well, I would say that the results have been sort of mixed. Uhuru
> has probably benefitted most African citizens, though the people
> living in towns have probably been helped more than those who live
> in the countryside. And in some regions, especially up in the
> Northeast, Uhuru hasn't made much difference at all. But if you
> are talking about Asian and European Kenyans, then I would say that
> in general Uhuru has hurt their economic condition, especially the
> Asians who have had to give up businesses. Though it is not the case
> that all Asian and European citizens have been hurt by Uhuru.

How is this person to respond if he is presented the closed rather than the
open question on Uhuru?

The closed question, sometimes called the fixed alternative question
has this one great disadvantage then. It forces the respondent to answer in
ways that may not accurately or at least not adequately represent his thinking
on the issue. The closed question loses flexibility.

The open question presents other difficulties. If the wording is
slightly ambiguous, different respondents will interpret the question in
different ways. The answers become so varied and unstandardised that analysis
is difficult. There are also the difficulties, and costs, of preparing the
open-ended material for analysis. Coding is the technical term for this process;
coding refers to the process of summarising open-ended material in terms of a
set of response categories. Coding open-ended material can be very time-consuming
and costly, especially if the sample is large. Take the response given to the
open-ended question on Uhuru cited above. If there were 1,500 such responses
the researcher would have to have a method of identifying and recording the
most important dimensions from each of these responses. And if one had, say
20 open-ended questions, there would be 20 x 1,500 or 30,000 paragraph responses
to prepare for analysis.

Despite the costs of coding the answers and the wide range of responses
produced, the open-ended question is widely used in survey research. It is
especially appropriate when the researcher is unsure what sort of information
a question will elicit and thus he does not want to foreclose possible responses.
Often then the researcher uses open-ended questions in a pilot or exploratory
survey with the intention of converting those questions into closed items in the
main survey. (A pilot survey is usually a small study carried out to test the
survey instrument and to further sharpen the research goals.)

Marjorie J. Mbilinyi, in a survey study of factors affecting the education of girls in Tanzania, provides an illustration. Mbilinyi carried out a pilot survey in which she relied almost exclusively on open-ended questions, such as:

Why do you think some parents do not send their children to school?

Can you think of any reasons why boys go to school?

What do you want to do when you grow up?

Mbilinyi comments on the shortcomings of such open-ended questions as follows:

So much depends on the training and ability of the researcher, acting as an interviewer, and on the verbosity of the subject. An interviewer can control the range of replies of his subject by consciously or unconsciously reinforcing certain answers or by literally 'feeding' the answers. /Open-ended questions 7 tend by their nature to be a problem for some subjects to answer. Our nonschoolgoing girls in particular responded with great difficulty to the more abstract questions such as, 'What reasons are there for boys to go to school?' (15, p.51)

These difficulties notwithstanding, reasons Mbilinyi, the open-ended questions are important in preliminary studies, for they enable the researcher to explore new problem areas in depth. In her research the answers to the open-ended questions were used to fashion a second interview questionnaire with fixed-alternative items. For example, she uses ranking questions whose alternative answers are based on the pilot survey. One of these asks subjects to rank in order of importance the following reasons for which some parents do not like to send their children to school:

1 -- misconduct of school children
2 -- children do not get secondary school places
3 -- school is foreign and has no value
4 -- children are needed to help at home
5 -- other .....................

Actually many surveys use a mixture of closed and open questions. This was the strategy adopted in a rural development survey carried out by the Institute of Development Research, Haile Sellassie I University. (10) The Ethiopian research team conducting this research wanted an extensive base-line survey which could provide data relevant to development projects for the Ghimbi Awareja of the Wollega Governorate in Ethiopia. Below are listed several different ways in which open and closed questions were used simul-taneously in this study.

Example    1 -- using an open-ended probe to give more depth to a closed item.

Do you believe that your standard of living will improve or decline in the next five years?

_____ I believe it will improve very much.

_____ I believe it will improve modestly.

_____ I believe it will not change.

_____ I believe it will decline.

_____ I can't tell.

Give reasons for your answer.

_____

_____

_____

Example    2 -- using an open-ended question to cover any alternatives not included in the alternatives presented in the closed question.

What problems do you have in your community?

____ delinquency

____ drugs (chat eating)

____ alcoholism

____ disability

____ crime

____ divorce

____ old age

____ none

____ others - specify

_____

_____

_____

Example    3 -- using an open-ended question to extend the material of the closed question.

In your opinion, has Ghimbi been given adequate attention by the central government?

_____ Yes, has been given adequate attention.

_____ Yes, has been given some attention.

_____ No, has not been given any attention.

_____ Don't know.

If reply is yes, indicate sphere of attention.

_____

_____

We have now reviewed three basic decisions which face the researcher as he begins to plan his survey: 1) whether to interview or use self-administered questionnaires; 2) whether to use an unstructured or structured approach; and 3) whether to use open or closed questions. The chart on the next page summarises our discussion and shows the interrelationships among the three decisions.

Basic Types of Survey Instruments

| | Unstructured interview | Structured interview or questionnaire, with generally open questions | Structured interview or questionnaire, with generally closed questions |
|---|---|---|---|
| Definition: | General list of topics to cover, but respondent is urged to talk about what- ever he considers impor- tant about the topics. Interview is not standard- ised across all the sample of respondents. | A specific and set list of questions is asked, but the respondent is free to use his own words in answering. Probes are used to illicit further materials. All respondents receive same questions and in the same order. | Similar to the structured questionnaire except that respondents are asked to choose their response from among the alternatives pre-selected by the researcher. |
| Illustrative Application Interview Studies: | An oral history study in which an old person is asked to tell everything he can recall, about a particular historical episode. You develop the interview to take advantage of the view- point he has of the episode. | Survey in rural development when you want the respondents to tell you what they consider the problems of the area to be. Useful when you are unsure what the possible range of answers might be. e.g., "What are the major problems facing the farmers of this area?" | Survey in rural development when you want to fix the range of answers permitted because you already know the possible answers e.g., "Which of the following fertilizers do you use?" |
| Self-administered Questionnaire: | not applicable | A study of citizenship in secondary schools where you ask students to write short essays to a series of such questions as: "What makes a man a good citizen?" "What are the main responsibilities a man has in his community?" | A study of citizenship in which you use questions such as "Do you agree or dis- agree that every student should serve in the national service?" |
| Basic Traits: | High flexibility In-depth material Low comparability | ← → | Low flexibility More superficial material High comparability |

CHAPTER 4

SURVEY TECHNIQUES: QUESTIONNAIRE CONSTRUCTION AND INTERVIEWING

INTRODUCTION

There is a large, and sometimes sophisticated, literature on techniques for writing survey questions and interviewing. There is no way that this literature can be summarised in a few pages, so this chapter is necessarily selective. It reviews some of the guidelines used by researchers who write questions and who train interviewers, but it does not review all that is known about these issues. The intent is to sensitise the reader to the complexity of the issues, and to remind the reader that every percentage or statistic or graph or table which appears in a survey research report is based on a questionnaire given to a sample.

The figures in a report are only as good as the questions used in the survey. Indeed, the figures and numbers and charts, no matter how mathematically sophisticated, come directly from the survey questions. A careful researcher will include in his research report the actual survey questions eliciting the responses he is analysing and on which he is basing his conclusions. The intelligent consumer of social science will pay close attention to these questions. He will want to judge for himself whether the questions were neutral or biased, and whether they provide an effective means of indexing the concepts used in the research report. The consumer of social survey research should develop a healthy skepticism. Just as he will want to know how carefully the sample was **selected,** he will want to know how carefully the survey instrument was constructed.

A prefatory comment is necessary before we turn to the guidelines which assist the researcher in writing questions and planning interviews. The guidelines we review primarily derive from American and European survey settings. They have to be translated <u>and</u> modified to be applicable in East Africa. The techniques of the public opinion pollster in the U.S. are not useless to the rural development researcher in Tanzania, but they are not directly applicable without modifications. And it is the local researcher who must make the appropriate modifications. He must do so in the context of the survey situation as he understands it.

It follows that there is an enormous burden placed on the survey researcher in East Africa. Everything the outside expert (including the author of this text) says about survey techniques must be examined for its relevance to East Africa. And only the researcher with knowledge of East Africa can undertake this examination.

## GUIDELINES

It has been emphasised that this chapter has to be selective. We proceed by considering a hypothetical case. A team of researchers in an East African country plans to conduct a base-line rural development survey. The sample for this survey has already been drawn. Thirteen hundred rural households from every part of the country have been selected. In each household, the head of the household will be interviewed. So also will every married woman of child-bearing age who lives in the household. The female interview will concentrate mainly on child rearing, children's health and birth control practices. The head of household interview will concentrate on agricultural issues and topics related to rural development.

This team of researchers includes many experienced persons who have taken rural surveys in their country. But no member of the team has ever directed a study as large as the present one, which eventually will employ 130 interviewers to collect data from an estimated 3,900 respondents (presuming an average of two married women of child-bearing age per household). The research team wants to proceed cautiously and correctly, and decides to ask a member of the National Statistics Office for some advice. They approach a person who has had experience in conducting large-scale national surveys and ask him to prepare them two memoranda. First, they request a memo on guidelines they should follow in constructing their questionnaires. Second, they request a memo on interviewing techniques which can be distributed to their interviewers during the training sessions which are planned. Here are the two memos he prepares; they will help you understand how the professional views questionnaire construction and interview planning.

### Guidelines on Questionnaire Construction

The most important rule is not to begin constructing a questionnaire by drafting questions. The team will end up with hundreds of separate questions, most of which will never be asked, and will risk overlooking important items.

The way to construct a questionnaire is to first list the specific research problems which concern you. Then ask, "What kinds of data will we need to study these problems?" And only then turn to the actual survey questions.

Here is an example of how to proceed:

Research Problem:   The nutritional standards of rural families, with particular reference to how nutrition affects children's health.

Type of Data
  Required:     a)  normal diet of the family
               b)  attitudes toward different kinds of foods
               c)  length of time infants are breast-fed
               d)  weight, height, age of every child
               e)  frequent children's diseases
               f)  etc.

Possible Survey   a)  "On the average, how many meals each week in this
 Questions:             household include the following foods?"

                       List of foods should include those known to have
                       high nutrition value as well as those known to
                       have low nutrition value.

                  b)  "Are there any foods which you or members of your
                       family refuse to eat because they are taboo or against
                       the customs of your people?"

                       Get list of such foods: use probes for specific foods,
                       such as "What about fish, does your family refuse to
                       eat fish?"

                  c)  "For how many months are infants generally given
                       mother's milk in this village?"

                       "Are there times when an infants might be given
                       mother's milk much longer than this?"

                       Find out why this might happen, and why some infants
                       may be breast-fed only for a short time.

                       Also, if respondent is a mother, **try to get the length
                       of time** she has breast-fed her own children and whether
                       a wet-nurse has been used.

By proceeding in this manner for every research problem of interest to the research team, you will insure that the data collected will relate directly to the major issues to be analysed.


Dummy Tables:  A more demanding way to proceed, but a more valuable one, is to create dummy tables. A dummy table anticipates how the data will actually be used in the analysis. It identifies exactly the variables to be analysed, and thus directs attention to the data necessary for the analysis. When you create dummy tables you will almost always ask relevant and pointed questions, for you are looking for the data which eventually will transform the dummy table into a table with actual numbers.

Return to the example just cited. The nutritionist on the research team will know that the mean height of different age groups can tell one a great deal about the nutritional standards of children. The following dummy table might be created:

### Mean Height of Boys and Girls

Mean Height

| Age Groups: | Boys | Girls |
|---|---|---|
| less than 12 mos. | _____ | _____ |
| 12 to 18 mos. | _____ | _____ |
| 18 to 24 mos. | _____ | _____ |
| etc. | | |

Creating this dummy table reminds the research team that they need a survey question indicating the sex of every child, indicating the age of every child in months, and indicating the height of every child. If this table is a very important one (as it would be to a nutritionist) then the research team may decide not to take the estimate of the parent regarding the height of each child, but to actually measure each child. Such a procedure is expensive in terms of research time, but gives much better data than asking parents, who often do not know the height of their children.

Creating dummy tables for all specific research problems is a useful way of establishing priorities for the questionnaire. Not every question of interest to every member of the research team can be included. Thus there should be some procedure for establishing priorities (all too often in team research the procedures are political rather than substantive; compromises are negotiated without sufficient attention to overall intellectual aims of the research). The dummy tables also alert researchers to the possibility that the same question or set of questions might serve more than one research purpose. The nutritionist wants to know if vegetables are a normal part of the household diet; the agricultural economist wants to know if vegetables are a standard crop. By creating dummy tables the nutritionist and the agricultural economist learn they need very similar data, and perhaps can devise one method for asking about vegetables which satisfies both research purposes.

The importance of thinking first about research problems, then about data needs, and only then about actual survey questions cannot be overemphasised. In a survey of this type, once the data are collected, it is too late go back and

ask a forgotten question.  What a researcher neglected to ask can never become
part of the analysis.  For instance, if a researcher asks a farmer how much
land he owns, but neglects to ask how much land he rents, the data will never
provide a complete picture of land use.  Hours spent thinking about what should
go into the questionnaire, before writing the questions, are never wasted hours.

## A Respondent-Centred Questionnaire

Often research staffs get carried away with the elegance of their
own thinking and draft questions which might be meaningful in terms of their
elaborate concepts but which are quite meaningless in terms of the experiences
of the respondent.  There are several ways to avoid this by making the question-
naire respondent-centred.

First, of course, is the length.  Very seldom can an interviewer
keep the attention of a respondent for more than an hour and a half.  And it
is dangerous to push a respondent to the limit of his attention span, because
he will start giving just any answers in order to bring the interview to a
close.  Of course the length of an interview depends on many factors.  A farmer
during harvest time might give fifteen minutes;the same farmer interviewed
between the harvest season and the next planting period might be willing to
spend a whole day answering questions.  Give the matter of the questionnaire's
length serious attention; it is better to get good, reliable data from an
hour's interview than to damage the whole project by trying for three- and four-
hour interviews.

The second matter to keep in mind is the language and vocabulary
which the respondent finds comfortable.  Write questions which reflect his
vocabulary, not that of the research team.  Try to avoid embarrassing questions,
though researchers around the world have found there is practically no limit
to the kinds of questions they can ask their respondents as long as they give
careful attention to how they word their questions.  If they are important to
the research goals, do not avoid issues of sex or politics or religion.  But
use care in how such issues are introduced and queried about.

The third and most important thing to keep in mind as one attempts to
make the questionnaire respondent-centred is the respondent.  As much as possible,
organise the questionnaire around the concrete realities of the respondent's life.

A farmer knows how many livestock he has, what kinds of seeds he uses, how
well his crops did last year, how easy it is to market his cash crops. Ask
him questions about such topics, and you will likely get reliable and
straightforward answers. However, the farmer does not know where he fits
into a systems model of communication or how high he is on a traditionalism-
modernisation scale. Beware of questions which are more closely tied to
complicated social science concepts than they are to the daily realities of
the farmer's experiences. (This does not mean one is prevented from studying
communication networks or the transition from traditional to modern farming
methods; it only means that these concepts must be indexed with respondent-
centred rather than researcher-centred questions.)

## Questionnaire Wording

There are dozens of guidelines for wording survey questions. Most
of these are common-sense ideas, and only a few of the more important guide-
lines are mentioned here. If the research team wants more extended commentary
on questionnaire wording a good treatment can be found in "Questionnaire
Construction and Interview Procedure," written by A. Kornhauser and P. Sheatsley
(12) or the well-known monograph by S. L. Payne, The Art of Asking Question (22).
The following are some illustrations of the type of guidelines found in these
textbooks.

     a) Avoid general words such as "often" or "many". When attempting to
        get quantitative responses, use concrete quantitative terms.

        Poor wording: "How much beer do you drink?"

        Better wording: "How many bottles of beer do you generally drink
                    each week?"

     b) Avoid questions which contain more than one question. Each item
        should seek one clear piece of information, and only one piece of
        information.

        Poor wording: "Are you a school leaver looking for work?"
                 (Respondent may be a school leaver not looking
                 for work, or looking for work but not a school
                 leaver.)

        Better wording: "What is the highest level of education you
                  reached?"

                 "Are you presently employed?"

                 IF NOT: "Are you presently seeking employment?"

c) Avoid technical terms, especially terms that might be familiar
to some respondents (perhaps the better educated ones) but not
others.

Poor wording:  "Do you have a high protein diet?"
Better wording:  "How many days a week do your meals include
                  meat, fish or chicken?"

d) Avoid words which can easily be misunderstood given the life-
situation of the respondent.  An interesting example of this
comes from a study conducted in the U.S. southern states among
very religious people.  These people were asked whether they
favoured 'government control of profits', and they all strongly
spoke against such controls.  This surprised the researcher until
he realised that the respondents thought they were being asked
whether they favoured 'government control of prophets'.

Poor wording:  "How frequently do you listen to foreign news on
                the radio?"  The phrase 'foreign news' is not part
                of a farmer's standard vocabulary, and he is likely
                to hear the question as if it asked about 'farm news'.
Better wording:  "How frequently do you listen to radio news about
                  what is going on in other African countries or in
                  other countries of the world?"

e) Avoid leading questions which tend to favour one rather than another
response.  A frequently made mistake is to write a survey question with
only one alternative presented to the respondent.

Poor wording:  "Some people in this region have told us that the D.C.
                is doing a poor job.  Would you agree or disagree that
                he is doing a poor job.
It is easy to agree with this question, because the respondent would
then be going along with the rest of the people in the region.

Better wording:  "Some people in this region have told us that the
                  D.C. is doing a poor job.  Others have told us
                  that he is doing a good job.  What is your view
                  on the kind of job the D.C. is doing?"

f) Avoid leading questions which associate important persons with one
part of the answer.  A higher percentage of persons would support
the second than with the first version of the following question.
"Do you support or oppose the proposal to have a high tax on luxury
goods?"

"Do you support or oppose President Kenyatta's proposal to have a
high tax on luxury goods?"

An even higher level of agreement would occur if the question included some negative stereotypes:

> "Do you support or oppose President Kenyatta's proposal to have a high tax on luxury goods which are mostly bought by rich people in Nairobi?"

g) Avoid vague or imprecise words, a particular danger when using open-ended questions. Vague terms produce a wide variety of responses and each respondent reads into the question his or her own meaning.

> Poor wording: "What do you think about health matters in this area?"

> Better wording: "Would you say that the health **services provided** in your village are adequate or inadequate?"

The second version directs the attention of the respondent to health services and to this particular village, replacing the vague reference to health matters and this area. It may be, however, that the intention is not to ask about health services but to ask about the health standards of the population. Thus a different directive is called for. Take care that in moving from **vague,** general questions to more precise questions you are moving in the direction of the data you want.

A huge number of such guidelines could be listed. Perhaps the best advice on questionnaire construction is to examine <u>every</u> item in the final questionnaire against a check-list such as the following one:

> Is the question necessary? Can you see how it will be used?

> Does the respondent have the information necessary to provide the kind of data you are seeking?

> Is the question respondent-centred?

> Is the question neutral or does it seem to favour a particular answer?

> Does the question contain imprecise or unclear or difficult words?

> Will the question be interpreted similarly by all respondents?

> Will the question make the respondent uncomfortable or suspicious or in any way lead him to resent the interview?

### Interviewing

The above memo on questionnaire construction was directed to a team of researchers preparing a national study of rural development in an East African country. We turn next to the art of interviewing, for it indeed is an art. And it is an art extremely difficult to teach. Some people seem to have a knack

for developing rapport with a respondent, and thus for getting the respondent's active cooperation in achieving the research goals. Other people are not so skilled; they are perhaps too aggressive in pursuing the interview and thus antagonise the respondent; or they are too friendly and accomodating, and thus fail to conduct the interview in a professional and neutral manner; or they are too hurried and cut the respondent off before he has really provided all the information necessary. And unfortunately, some interviewers are simply dishonest. They sit at home or in a local bar filling out one questionnaire after another in a practically random way. (Effective supervision of field staff should detect the dishonest interviewer, but not all research studies include effective field supervision.)

There is no agreed upon technique for making a good interviewer. Of course when the researcher is doing all of his own interviewing, the success of the project depends largely upon his skill. But when, as in our present illustrative case, a large survey study depends on many hired interviewers, then the research staff has to concern itself with interviewer training.

Usually this training has two main parts. First, a thorough and detailed introduction to the questionnaire, with the research purpose behind every question explained by the research staff. In a well-organised survey there is often an interviewer manual which instructs the interviewer on how to ask complex questions and how to record the answers. The rural development survey in Ghimbi, Ethiopia, mentioned earlier, produced such a manual. Here are **four questions from that survey,** followed by the instructions which appear in the interviewer manual:

Qu. 4. Perception of Resource Stability and Fertility.

A. What is the nature of soil fertility in your area? Break down your total holding on the basis of the following:

    1. very fertile         _____
    2. moderately fertile   _____
    3. very low in fertility _____

B. How do you rate the soil stability condition in your farm?

    1. not eroded          _____
    2. moderately eroded  _____
    3. extremely eroded   _____

C. How do you think the fertility status of your soil could be improved?

    1. Fallowing           _____
    2. Terracing           _____
    3. Fertilizing         _____
    4. Rotating crops     _____
    5. Not possible to improve _____

D.  What is the condition of the natural vegetation in your area?

    1.  Stable      _____
    2.  Depleting   _____
    3.  Depleted    _____

From interviewer's manual:

Q.4.    Perception of Resource Stability and Fertility

Q.4.    a.  Nature of Fertility of Soil
            This question attempts to find out the fertility condition
        of the holding - of each household.  Ask the interviewee what
        fraction of his land is very fertile, moderately fertile and very
        low in fertility.  Enter in the blank as accurately  as possible
        the ratio of the land under each category of fertility.

Q.4.    b.  Rating of Soil Stability Condition
            For this question, record the general condition of the
        holding of the household in terms of degrees of erosion.
        Select only one of the degrees of erosion which best
        describes the holding.

Q.4.    c.  Improvement of Soil Fertility Status
            Record what, according to the interviewee, are the
        methods   for improving the soil fertility status of the
        holding by ranking the methods of improvement in order
        of preference by the interviewee.

Q.4.    d.  Condition of Natural Vegetation
            Explain to the interviewee that you are interested in the
        general condition of the natural vegetation of his surrounding.
        Attempt to find out which of the categories best describes the
        condition of the natural vegetation in the area.  By "stable"
        it is meant that the natural vegetation is in a state of virtual
        completeness or intact.  By "depleting" it is meant that the
        actual vegetation is being used at an accelerated rate without
        replacement.  By "depleted" it is meant that the natural
        vegetation is almost gone. (10)

A manual of this type would be carried with the interviewer and used during
the field work as a check to insure that all the necessary information is
being collected and correctly recorded.

        The second major aspect of interviewer training is practice inter-
viewing.  The more the better, and the practice sessions should when possible
include a trained interviewer who can spot and quickly correct faulty inter-
viewing techniques.

        Below is a memo prepared by the consultant to the rural development
research team, drafted as if it were to be distributed to the interviewers during
their training session.

## Guidelines on Interviewing

The interviewer plays a critical role in survey research. It is the data he collects and records which become the basis for analysis. The present study is designed to provide important base-line information on rural conditions across the nation. This information will be used to prepare development projects in health and sanitation, in marketing and production, in education and training, and many other aspects of rural development. Faulty and incomplete information may well lead to faulty and useless development projects.

An interview is a social exchange between the interviewer and the respondent. Insight, intuition and personal rapport play a part in the outcome of this exchange, just as they do in any other social situation. A good interviewer realises that the interviewee is guessing at the motives of the interviewer and is perhaps tailoring his answers accordingly. Thus it is very important that the interviewer realise how his own behaviour and dress and manner might affect the interview situation. As much as possible the interviewer must minimise those aspects of self which might detract from the goal of obtaining reliable information.

Gaining Cooperation: Many persons like to be interviewed. They are complimented that someone is interested enough in their opinions and their lives to have sought them out. For this reason, gaining cooperation is not usually a difficult task. Yet it is also the case that interviewing takes time, and for people busy on their farms and in their occupations time is valuable. It is for this reason that sometimes cooperation is more difficult to obtain.

When persons are reluctant to be interviewed you should point out to them the importance of the project. Stress that the information from this survey will be important in planning rural development projects.

A person reluctant to be interviewed will often remark that his neighbor or his relative is not so busy, so why not interview him instead. At this point it is very, very important that you stress that he has been especially selected. You will of course not be able to give a lesson in statistical sampling, but you can explain that the study depends on getting full cooperation from all the families especially selected. Tell him that he was not selected by the village headman, but by the Central Bureau of Statistics.

The most important thing in getting cooperation is to be honest with
the respondent. If he asks you what kinds of questions you are going to ask,
tell him in a general way the topics to be covered. If he asks how long it
will take, again tell him how long an average interview lasts. If he asks
what is going to happen to his answers, explain that they will be put together
with those from 1,300 other questionnaires. It is the **answers** from all over
the country which will be analysed together. Make sure you emphasise that no
government official, especially the headman, will see his specific answers.

Interview Setting: The ideal interview setting is a private place where just
you and the respondent can conduct the complete interview. This is not always
easy to achieve, so we point out two guidelines.

a) Try to avoid multiple-sessions. Complete the interview in the first
   session. If the respondent says about half-way through that he has
   to go elsewhere, but will finish later in the day or the next day, urge
   him to finish now if at all possible. It is better to break the
   interview into two sessions than not get a complete interview, but
   it is best to complete it in one session.

b) Try to avoid public interviews, that is, interviews in which persons
   other than the respondent are present. Again, this will not always
   be possible, but keep in mind that respondents often answer
   differently in public than in private. Also, if other persons are
   present, there will often be arguments about what is the correct answer.
   As much as possible, take the answer from the person who is the principal
   respondent. This will standardise the questionnaires from one house-
   hold to another.

Coverage: The questionnaires you will be administering are fairly complicated,
in that some of the questions will be appropriate for some households but not
others. And some of the questions will be appropriate for some of the people
in the household, but not all of the people. The most important rule of cover-
age is that for any given respondent _every_ question appropriate to him should
be asked. Here are a few specific guidelines which will help.

a) The questionnaire is a fixed-order questionnaire. You should ask the
   questions in the sequence in which they appear during each interview.

b) Don't skip questions.  Sometimes you will think that you know what
   the respondent is going to answer and you will be tempted to skip the
   question and answer for him.  This is very bad interview practice.
   It is important to get the respondent's specific answer, not your guess
   as to what he is probably going to say.

c) Keep the respondent to the topic.  This is a most critical rule.  Some
   respondents will start to tell a long story in response to a question,
   telling you things that happened to his family several years ago.
   When this happens you must try to be firm.  Because if he spends
   fifteen minutes answering a question which should take only one
   minute, he will tire of being interviewed long before you are finished
   with the questionnaire.

   Sometimes the respondent will start to answer questions which you have
   not gotten to yet (remember, he does not know what is in the question-
   naire).  At this point you can firmly point out that you are interested
   in his information, but that you will be getting to that later.

Misinformation:  Every time an interviewer records misinformation or unreliable
information, the research project is damaged.  There is no way during the analysis
to sort out poor quality from good quality information, thus every mistake made
during an interview reoccurs whenever the question on which the mistake was made
is used in analysis.  The entire research project is only as good as the infor-
mation you record during your interview.  Poor quality information can be caused
by many different factors, not all of which of course would be the fault of the
interviewer.  But it is true that the interview stage can cause more damage to
the quality of information than any other part of the research.

What is poor quality information?  It is information which is incorrect
or is incomplete.  If the head of the household says that there is a television
set in the house, but in fact there is not, incorrect information has been
collected.  If there are two children under six in the household, but the health
record of only one of them is collected, that is incomplete information.  In-
correct and incomplete information are equally damaging.

Sometimes the respondent will deliberately mislead you.

a) He may actually lie, in order to cover up something he does not want
   you to learn about.  But in fact this type of misinformation occurs
   rarely, and in any case is very difficult to correct.

b) More often, and therefore more of a problem, is the respondent who tells you what he thinks you want to hear. He wants to please you, and thus biases his answer. This is why it is important for you to be an interested, but neutral listener. Do not indicate that you think one answer is more correct or better than another answer, for if you do you will often get that answer even if it is not correct from the viewpoint of the respondent. **Never** express displeasure at an answer given by a respondent. For example, if he thinks you disapprove of how the household food is prepared, he will begin to tell you that it is prepared in a different and more acceptable way.

c) Even when the respondent does not worry about pleasing you, he may wish to use you as a means of speaking indirectly to authorities. This is a major source of misinformation in survey studies. The questionnaire becomes not a reflection of the true conditions but a reflection of the conditions the respondent wants the authorities to think are true. For instance, if he fears the authorities, he will hide information. More often is the case where he wants to demonstrate the need for more government services. For example, perhaps an agricultural extension officer visited him three times last year. He does not consider that enough, and thus **tells** you that he has never been visited so that more officers will be sent to the village. Again, this is a cause of misinformation which is difficult to solve. But by being alert, you can try to insure that you are given correct information.

d) A further **cause** of misinformation is ambiguity in the question or confusion in how you ask it. The question-writers have tried to eliminate all ambiguity, though it is difficult to be completely success-ful. Confusion in how a question is presented is of course the fault of the interviewer. We have so many practice sessions in order to find the parts of the questionnaire likely to cause confusion, and to try to make appropriate corrections before we go into the field.
There is one rule to keep in mind. If you find that the respondents seem not to understand a particular question, make certain that you explain it in the same way each time. Do not change your mind about what a question means sometime during the research. Consult your inter-viewer manual; and if a problem persists with a particular question, consult your supervisor. Do not modify a question or alter the order of the questions asked unless specifically instructed to do so.

Questionnaire Editing: Each interviewer should review his or her questionnaires at the end of the day. This check, or editing as it is often called, should accomplish three things. First, check and double-check that the identification numbers have been correctly entered into each appropriate place on the questionnaire. Check to insure that the identification numbers for the head of household questionnaire and the supplementary questionnaires used with the married women correspond.

Second, watch for misleading marks on the questionnaire, marks of the kind which might confuse a person trying to code the questionnaires several months from now. For example, sometime you will place a tick in one box only to find that the respondent changes his mind about how he wants to answer. When you tick his new answer it is easy to forget to erase the initial tick. This will lead to confusion at a later stage, unless you carefully review every page of the questionnaire.

Third, review what you have recorded in open-ended questions. Your writing must be clear enough for anyone to read it, even someone far removed from the interview situation and who will be coding the material several months from now. Only by checking the questionnaire soon after the interview will you remember things that you failed to record at the time.

## Conclusions

The two memos from a consultant to a research team illustrate many of the principles which can lead to sound questionnaire construction and reliable interviewing practice. Yet as previously emphasised, the review in this chapter is highly selective. There are numerous complicated issues in the technique of survey research which have not been reviewed. For instance, nothing has been said about language and translation difficulties when the sample includes more than one linguistic group, though this is a serious issue for **the survey researcher in East Africa. Moreover, we have not dealt with the way in which differing cultural contexts can markedly affect the reliability of survey items. Among certain cattle-owning tribes of East Africa, for instance, it is an insult for a stranger to inquire how many cattle are owned, or to ask about family size. The naive survey researcher unfamiliar with cultural traditions and taboos is** not likely to obtain reliable or valid information.

It is beyond the scope of this text to deal adequately with the enormous range of problems which influence the accuracy of survey data. The present task is simply to help the reader understand that behind an effective survey is careful planning, creative questionnaire construction and detailed attention to interview techniques. A survey which does not exhibit these characteristics should be approached with caution, even skepticism.

Unfortunately many poor surveys are conducted. This has especially been the case in the early years of survey work in East Africa. Several factors contributed to poor quality work, including most strikingly the large number of surveys carried out by Westerners not sufficiently familiar with cultural and linguistic conditions in the East African countries. Poor survey work, however, is not all to be laid at the feet of researchers from the U.S. or Europe or England. Local researchers have not always applied high standards either.

We mention poor quality survey work in East Africa not to identify particular studies or to criticise particular researchers. Our purpose is of a very different kind, and derives from a protective instinct toward survey research. Some sensitive observers of social science in East Africa, observers who can be found in the universities (sometimes in social science departments) and in the governments, have not overlooked the many dubious pieces of survey research which have been reported in their countries. These observers have at times been led to the conclusion that survey research itself is a dubious proposition. Often they reason that a technology so clearly Western in its origins and in its basic principles has no part to play in third world countries.

This, we believe, is a hasty conclusion. The fact that survey research does not always measure up to professional standards does not mean that there are no standards. This chapter has attempted to make clear that there are standards against which survey questionnaires and interviewing practices can be measured. The role of critics is to demand of researchers that they meet exacting standards. Anyone familiar with East African social science research will quickly see that when these standards are met, the results of the research can be very valuable indeed.

CHAPTER 5

NON-SURVEY DATA:  THE PUBLIC RECORD AND OBSERVATION

There are ways to collect systematic social data other than through
the sample survey.  In a previous chapter the distinction was made between
reactive and non-reactive research, a distinction related to whether or not
the researcher deliberately provoked a response from the persons being
studied.  The chief example of reactive research is the sample survey; the
chief example of non-reactive research is the use of the public record.  The
latter type of research will interest us in this chapter.  In addition, we
include a brief discussion of another research technique - systematic obser-
vation.

We start by reviewing two points.  First, although sampling theory
is commonly associated with survey data, it is also applied to other data
collection methods.  The researcher using public documents or observational
techniques will frequently make use of sampling theory in order to generalise
from a limited set of observations to a larger population.  Whatever the source
of data, sampling comes into play whenever the researcher wants to generalise.

Second, like sampling theory, statistical analysis is generally used
with survey data, but is by no means limited to such data.  If data have been
systematically collected and if they are quantifiable, the statistical tools
outlined in the following chapters will help discover and efficiently summarise
relationships in the data.

In reviewing non-survey data sources we will be less detailed than
we were in commenting on survey data.  It would greatly extend this manual
if all the characteristics of public record research or of observational
research were described.  Our purpose in this chapter is to acquaint the
reader with non-survey data, but not answer all possible questions about how
to collect and use such data.

The Public Record

Contemporary societies collect and report an enormous amount of data
about themselves.  These data, in the hands of imaginative social scientists,
are invaluable aids in the study of a society.  The primary sources of these

social data are the governments themselves especially through the annual
reports of ministries. Then there are also the newspapers which supply a
continuous account of relevant social, political and economic happenings.
In addition to government reports and newspapers are the publications of
particular institutions in each society. A national council of churches
is often a source of information about religious practices and trends in the
population. Individual businesses as well as industrial associations and
councils can be a rich source of economic data.

In a moment we turn to some illustrations of public record research,
First we need to emphasise why these data sources must be approached with
caution. There are three main points the social scientist will **keep** in mind
as he attempts to study society through the use of public documents.

The Institutional Bias: Nearly all public records are the product of self-
reporting. The Uganda Ministry of Education collects and distributes edu-
cational statistics; TANU makes available candidate profiles prior to an
election in Tanzania; the East African Automobile Association publishes data
about auto travel: the evaluation of the Special Rural Development Project
in Kenya comes from the ministry and the research institute (I.D.S.) most
directly involved in the project.

Self-reports are often biased. They exaggerate facts which work
to the credit of the reporting agencies and underplay facts which would dis-
credit it. Crime statistics are a particularly notorious example of this
bias. When a police department is trying to obtain a budget increase it might
report crime statistics in a way suggesting an increase in the crime rate.
Perhaps it reports every loss of property as possible theft. Another police
department, which has just had a large budget increase and now wishes to
justify it, might produce statistics showing a decrease in crime. This
department reports only those cases where theft is clearly demonstrated.

Institutional biases by reporting agencies are easily explained.
To stay in business an agency attempts to demonstrate that it is doing what
it is expected to do, and that it can do even better if its budget is
increased. Churches seeking support from mission societies are as tempted
to play with the statistics as are departments trying to get money from the
central university budget or ministries trying to protect themselves against
parliamentary attack.

However understandable from a political or economic point of view,
the institutional bias creates problems for the researcher working with
public documents. A careful researcher will seek out possible sources of
distortion in the public record and will take these sources into account.
And the careful reader of research reports will take similar cautions. He
will check to see if the writer of the research report acknowledges possible
sources of error; and he will want to know if the writer has attempted to
correct for them.

The Durable Artifact: Have you ever wondered why archeologists can describe
in detail the cultures of ancient Greece and yet provide such limited descrip-
tions of the cultures of ancient Europe or Africa? Archeologists reconstruct
how preliterate peoples lived by studying the things they left behind --
monuments, temples, tools, weapons, ritual objects. These are the material
artifacts of a culture. Some types of artifacts resist the corrosive effects
of time more successfully than others, and this in large part explains our
uneven understanding of different ancient cultures. The settled, agricultural
people of ancient Greece made their artifacts of stone or metal and thus left
behind a much more complete record of their way of life than did the more
nomadic people of Europe or Africa who made their artifacts of animal skin or
woven grass for easy transport. The culture of a nomadic people may be every
bit as complicated and sophisticated as that of a settled people, but it is
not as likely to be expressed through durable artifacts.

This idea of the durable artifact is today relevant to how the
social scientist approaches the records made available by governments and
other contemporary institutions. Some records are more durable than others.
The political historian in Kenya, for instance, is well aware of this fact.
Say that he wishes to compare the colonial administration of agricultural
development with the colonial administration of political security from 1940
to 1960. The former set of records would be much more complete than the latter
set. Prior to independence, the colonial government systematically destroyed
many records about how they attempted to control African independence movements.
And these political movements seldom kept written records of their own
activities and memberships for fear that such records would fall into the hands
of the colonial police authorities. Thus the historian trying to reconstruct
political opposition movements and their suppression has a much more difficult
time than when he is trying to reconstruct colonial agricultural practices.

The durable **artifact** principle has many applications. Those things
which powerful institutions in society consider important are those things
which receive the attention of record-keepers. Governments in nearly all
societies keep elaborate economic records; they keep much less detailed
health records. Thus the social scientist can better determine what the tax
burden of a district is than he can learn whether the children of that district
are well-fed. The United States, despite being a very record-keeping nation,
has no reliable national statistics on malnutrition, for instance. Yet it has
very extensive data on communication and transportation networks, information
of importance to the industrial and commercial sector as it plans production
and marketing systems.

A similar complaint about Kenya was recently voiced by a team of
researchers; this group, in preparing a report popularly known as the I.L.O.
report, had reason to look carefully at the statistical sources of Kenya.
They note that areas of social life are very unevenly reflected in these
statistics.

> A good example is the informal sector. We have indicated its
> importance in several chapters, but we have had great difficulty
> in obtaining any clues at all to its scale and composition.... (9, p.587)

> Similar points can be raised about subsistence agriculture.
> One reason for the neglect of this sector may be the paucity
> of data about it.... Yet it is the starting point of a socially
> oriented development plan. (9, p.588)

> Such /educational/ statistics as are available are significantly
> more complete for formal than for informal education, and for
> maintained than for unassisted institutions.... Published
> statistics tend to concentrate too narrowly on enrolments,
> projected enrolments, and pupil-teacher ratios, to the neglect
> of other vital information. (9, p.593)

> There are **virtually** no published data in the whole field of
> training on the effort of private firms and voluntary agencies. (9, p.593)

Secrecy: Often there are records which simply are unavailable to the social
researcher. In Kenya, for instance, it is difficult to obtain detailed
information on land transactions: who is buying what plots of land and for how
much money. Similarly with the distribution of businesses formerly held by
non-citizens: to whom are those businesses going and for what prices? Only
incomplete information is available on both of these issues.

Another major area of secrecy concerns military matters. Though
understandable, this secrecy nevertheless imposes a burden on the researcher.
It is nearly impossible to find out the exact proportion of the budget going
to the armed forces and to intelligence operations, for instance. And this
is true in every country.

Government secrecy extends beyond military and security matters.
Although court decisions are announced publicly, the deliberations of the
justices take place in private. And although the parliament debates are
recorded in the Hansard and in the newspapers, not all the important legisla-
tive business takes place on the floor of parliament. That which occurs in
committees or party caucuses is not usually recorded or made available for
public inspection.

In short, it is much easier to get information on the finished
product of a parliament or executive or judicial branch of government than
it is to get information on the process out of which that product emerged.
And this holds for other important institutions. We can learn from the
public record that the University of Nairobi admitted 30 per cent more students
than the year before; we cannot easily learn about the considerations which
went into this decision.

Biased reporting, incomplete records, and secrecy are major problems
which plague the researcher who depends on the public record for his data.
But these problems are not insurmountable. Some excellent social research has
been reported which draws heavily on the public record.

Sometimes this research is known as aggregate statistical analysis.
The term aggregate statistics refers to percentages or means or other statistics
which summarise the traits of a large collection (an aggregate) of people,
usually entire populations. It is in government records that the researcher
most often finds aggregate statistics, though the use of these statistics is
subject to the cautions noted above. For example, the following aggregate
statistics come from government sources of East African countries:

> per cent of population with different levels of schooling
> per cent of population taxed at different levels
> per cent increase/decrease of specific crimes from one year to next
> male-female ratio of different age groups
> mean savings rate for different regions

Extensive use of aggregate statistics appears in the I.L.O. report,
Employment, Incomes and Equality. A major conclusion of this report is that
there is a need to redress the regional inequalities in income and essential
services across Kenya. The identification of regional inequalities depends
heavily on analysis of aggregate statistics from the public record, as is

illustrated in the following table:

### Distribution of Social Services by Province, 1970

| Province | % of total pop.1969 | % of school enrolment, 1970 | | % of NHC[1] housing expenditure 1970 | Number of people | |
|---|---|---|---|---|---|---|
| | | Primary | Secondary | | per hospital bed | per medical practitioner |
| Rift Valley | 20.4 | 14.7 | 12.1 | 6.0 | 820 | 1,755 |
| Nyanza | 19.4 | 16.1 | 13.1 | 1.2 | 1,269 | 2,219 |
| Eastern | 17.4 | 20.2 | 13.6 | 2.4 | 834 | 1,734 |
| Central | 15.3 | 24.9 | 22.9 | 15.1 | 766 | 1,287 |
| Western | 12.3 | 13.1 | 10.1 | 2.9 | 1,033 | 3,569 |
| Coast | 8.6 | 6.3 | 9.3 | 7.2 | 511 | 707 |
| Nairobi | 4.4 | 4.4 | 18.7 | 65.2 | 152 | 84 |
| North-Eastern | 2.2 | 0.3 | 0.2 | - | 1,308 | 1,230 |
| Whole country | 100.0 | 100.0 | 100.0 | 100.0 | 715 | 871 |

[1]National Housing Corporation.

Source: Employment, Incomes and Equality (Geneva: International Labour Office, 1972), p.301.

In using aggregate statistics the researcher cannot determine the trait of an individual, but only of an aggregate of individuals. This is the major distinction between aggregate statistics and survey data. A study of examination performance using aggregate statistics tells the researcher the average performance of a school but not the specific performance of individuals in that school. Additional data are necessary if the researcher wishes to study examination performance at the individual level.

## Examples of Public Record Research

In a very elaborate study, Nicholas Nyangira has investigated the relationship between relative modernisation and different public policies in fourteen districts of Kenya. Nyangira's research is far too extensive to be adequately summarised here; but we can use it to good purpose by showing his use of the public record. The initial task faced by Dr. Nyangira was to provide some concrete indicators of his concepts. We restrict ourselves to the concept of modernisation, which Nyangira defines in **terms** of urbanisation, communication, transportation, education and so forth.

His actual modernisation variables are thirteen, as follows:

1. Number of registered cooperative societies
2. Number of agricultural extension staff
3. Population per square kilometer
4. Population size
5. Number of students attending secondary schools
6. Number of pupils in primary schools
7. Number of students passing School Certificate and East African Certificate of Education
8. Livestock products sold in K£'000
9. Income from coffee in K£'000
10. Number of post office rental boxes
11. Number of branches of banks
12. Number of telephones
13. Size of Asian population

The sources listed by Nyangira for his measures include the following, all save one of which are standard public record documents:

> Kenya, Department of Co-operative Development, Annual Report, 1962 (Nairobi, Government Printer, 1964); J. Heyer, D. Ireri, and J. Moris, Rural Development in Kenya (Nairobi, Institute for Development Studies, 1969); Kenya, Population Census, 1969, Vol. I (Nairobi, Government Printer, 1970); Kenya, Ministry of Education, Annual Report, 1969, (Nairobi, Government Printer, 1969); Kenya, Department of Agriculture, Annual Report, 1966/67, Vol. I (Nairobi, Government Printer, 1969); Kenya Post Office Directory (Nairobi, Kenya Litho Ltd., June 1970); Kenya Telephone Directory (Nairobi, Kenya Litho Ltd., 1971). (19, p.22)

What we have in Nyangira's work, then, is an extensive use of the public record to investigate a major topic in the field of development adminis tration. The regular publications of the Kenya Government become a source of the variables which index modernisation, and this concept in turn can be related to the kind of public policies we find in the different districts of Kenya.

In a paper titled Land Transaction in Kiambu (11), James G. Karuga provides another illustration of intelligent public record research. Dr. Karuga, an economist, wished to examine whether the value of land, as is often hypothesised declines with distance from a central market. Moreover, does the volume of land transactions decrease as one moves further away from a central market? His research site was the Kiambu District of the Central Province in Kenya. The source of data was the official Land Registry, from which he was able to obtain information about whether a given Farm Parcel had changed hands between 1956 and 1971.

Karuga's comments on his methodology are instructive, for they reveal some of the difficulties in public record research and some of the ways in which the researcher, conscious of the need to avoid self-deception, can correct for

these difficulties. He finds from the Land Registry information that 30 per
cent of the farms have undergone some transaction. He reasons, however, that
this proportion gives a false picture of the amount of social and economic
change in Kiambu. Based on extensive field research using the survey methodology
in one community of the Kiambu area, Karuga estimates that closer to 45 per
cent of the farms have been involved in transactions. Karuga notes that the
undercounting in the public record derives from the fact that many farm sales
and purchases occur within the clan or family and are never recorded.

It might be thought that this distortion would invalidate his research.
It would, if Karuga were primarily interested in estimating the frequency of land
transactions. But this is not his research goal; what he wishes to determine is
whether the frequency of transactions varies depending on how far the area is
from the central market place (Nairobi, in his research). If we can assume that
the extent of undercounting in the public record is similar from one area to
another then the data can still be used. To understand this point we need to
appreciate the difference between an accurate ranking of the units of analysis
and an accurate measurement of the units. The Land Registry does not accurately
measure how many transactions occur; but it can still accurately rank the different
communities in the Kiambu area. This can be shown as follows:

|  |  | Per cent of land transactions according to the Land Registry | The "true" per cent of land transactions |
|---|---|---|---|
| Community | 1 | 40% | 55% |
| Community | 2 | 33% | 46% |
| Community | 3 | 21% | 38% |

As long as the communities are placed in the correct ranking by the methodology
used, many statistical procedures will work perfectly well. And in the example
above Community 1 has more transactions than Community 2, both in fact and
in the data available to the researcher.

To obtain the true per cent would have involved a much more expensive
data collection operation than that chosen by Karuga. In fact, detailed survey
work in every area was prohibitively expensive. As the author writes, "Our
approach, however, enables us to have an overview of the processes over a wider
area -- the whole district -- rather than a single locality."

This research illustrates another difficulty with public record data,
and again Karuga's comments indicate that he is sensitive to these difficulties.
The Land Registry records what happens to land parcels; it does not record
anything about the persons involved in land transactions. As Karuga writes,
"One of the drawbacks of our approach of using the farm as the unit of our
analysis is that we cannot know who is involved in these processes of purchases--
i.e., their previous land holdings, income, district or location of origin,
and the reasons for their purchases." Karuga is here noting a chief weakness
of public record data; such data are amenable to some but not all research
questions. As we said above, the researcher has no control over the stimulus
presented to the respondent. In the present instance, the researcher could
not determine answers to a series of important questions about Kiambu land
transactions. (11, p.6)

Again, this does not invalidate the research. The data collected
were satisfactory to the research question: do land transactions decrease in
frequency as you move away from a central market place? That the data could
not be used to answer other research questions is no cause for dismissing them.

The political scientist is favoured over other social scientists when
it comes to public record information. The reason is obvious. Governments
keep records about their activities and their personnel. Thus many of the
things of interest to the political scientist are frequently and extensively
described in the public record.

This has been especially true of elections. Perhaps no field of politi-
cal study has so effectively used public record data as have election studies.
A good case of this is the recently completed study of Tanzania's 1970 national
elections, a sophisticated and important research project which drew heavily
on public record materials.

Let us take just one example: the issue of candidate selection.
TANU was faced with the problem of selecting 240 candidates from the more than
1,100 persons who presented themselves as possible candidates for Parliament.
Bismarck U. Mwansasu, using only public record information, sheds a good deal
of light on the criteria applied by TANU in the various stages of this selection
process. He then compares the TANU criteria with the final election results,
and thus  can contrast the voter choices with the preference orderings suggested

by the party. Although the results of this research are fascinating, a detailed
review would distract us from our immediate concern with methodology. What
from the public record was available to Mwansasu as he set about his research?
Mwansasu writes:

> A prospective candidate is required to fill in a form which asks
> him to provide his name, date and place of birth as well as that
> of his parents and his occupation. A candidate who is a farmer
> or peasant is further asked if he lives in an ujamaa village. If
> he does not, he is asked to explain the reasons; if he does, he
> is asked to name it, its location, its chairman, the date it was
> formed and when the candidate joined it.... Secondly, the candidate
> is asked to give information on his background as a Party member....
> Thirdly, he is asked to give a brief account of his career mentioning
> the various positions held and the period during which they were held.
> ... Fourthly, he is asked to provide information indicating whether
> he has been convicted of any crime in a court of law.... Fifthly,
> he is asked to indicate whether he has made any attempts to contest
> any Party offices which were rejected by TANU.... Sixthly, he is
> asked to state whether he has attended adult education classes, any
> seminar on political education or any training in national defence
> such as the National Service. Finally, the candidate is asked to
> give an autobiographical account. (18, pp. 141-2)

This is very rich political information indeed, and its reliability
is tested when the prospective candidate faces questioners at the Annual Dist-
rict Conferences. Mwansasu is well aware of how useful these data are.
"Because every prospective candidate throughout the country fills in the same
form the information is extremely valuable. Not only is it the only official
source on the candidates' background but the fact that it has been attested
on oath as correct and accurate suggest that it is also reliable." (18, p.142)

The data are used to ask such questions as: What types of people
offer themselves for recruitment into the country's leadership roles? What
criteria are used by the District Conferences in narrowing the number of
potential candidates? What criteria are used by the National Executive Committee
of TANU in choosing the final candidates? How do the candidates finally chosen
by the voters differ from the pool of those who initially presented themselves?
To report just one finding, 74 incumbent parliamentarians initially presented
themselves as candidates; 58 of these were given a preferential ranking by the
District Conferences, all of whom were subsequently accepted by the National
Executive Committee as final candidates. The voters then rejected 22 of these
incumbents, a high proportion compared to the number rejected from other
occupational groups. In drawing conclusions about the high rejection of in-
cumbent parliamentarians, Mwansasu reviews the problems associated with consoli-
dating democratic socialism in Tanzania. His study, thus, illustrates how
extremely important social issues can be investigated with information from the
public record.

Researchers who use the public record as a source of social data
often combine the data collected with information obtained from other sources.
An excellent example of combining different data collection techniques is found
in an extended study of Kenya secondary school leavers. This study, known as
the Tracer Project, is directed by H.C.A. Somerset and Kabiru Kinyanjui. Its
aim is "to find out what happens to $\overline{/}$ secondary school leavers$\overline{/}$ after they leave
secondary school, and to find the relationship between this education and
after-school activities".

The researchers traced a national random stratified sample of leavers
from 22 Kenyan schools which had Form IV classes in 1968, a sample consistin
of 3,179 African respondents. In a publication from this study Mr. Kinyanjui
describes the data collection strategy; a description which is here rep: duced
at length in order to reveal the ingeneous combination of public record data,
informant data and systematic interviewing:

> The data collected in this study are of two kinds. There are firstly
> East African Certificate of Education (E.A.C.E.) results showing the
> performance of each secondary school leaver in the sample. These
> results show scores in each subject, and the overall performance in the
> candidate's best six subjects which determine the division of the
> certificate awarded. These results were obtained from the examination
> section of the Ministry of Education. The second kind of data is
> related to what h.ppened to each school leaver after completing Form
> IV. This information was obtained by tracing school leavers retrospec-
> tively. A number of strategies were adopted to obtain this information:
>
> 1. School records
>
> In some schools career masters and mistresses keep records of what
> happens to their former students. However, these records are often
> neither complete nor systematic. The task of keeping up-to-date
> ...cords of the whereabouts of school leavers is difficult where there
> are .. equent changes of staff, and the cost in time and money is beyond
> the means of many schools. Where records existed, they proved
> extremely useful by giving us a starting point for carrying out further
> work.
>
> 2. Contacts with former students
>
> These contacts were made by using information from schools, especially
> from career masters and mistresses. Where information of this kind
> was lacking, we consulted major national institutions. This gave us
> opportunities of meeting students from our sample schools who were
> studying in places like Kenya Science Teachers College, Kenyatta
> Medical Training Centre and the Teacher Training Division of
> Kenyatta College. After initial contacts were made at these places,
> we began tracing leavers in some of the major urban centres.

On the basis of information obtained from our primary sources,
various methods were used to supplement or correct this information
and to obtain completely new information.  In Nairobi, it was possible
to make contacts on the telephone or to arrange to meet the respondents
after office hours or at their places of work.

Our interviewers, mostly university students, contacted respondents
at their places of work in the other major urban centres of Mombasa,
Kisumu and Nakuru.  Although the number of people contacted in this
way was small, it provided us with an opportunity to interview the
leavers and have them describe their job experiences.  The benefit
of these interviews was that school leavers could personally correct
or confirm the information we had obtained from other sources.  In
this way we were able to evaluate the accuracy of our sources.  We
found it difficult to interview school leavers working in rural areas,
but we attempted this, especially when we visited schools in the
Eastern and Central Provinces of Kenya.

Data were collected from those who had continued with further
education or training, from those who had gone into direct employment,
and from those who were unemployed.  These data showed the activities
of each school leaver for every year he was in the labour market.

Studying school leavers retrospectively has its problems, one of
which is getting systematic information on an individual's activities
over a period of time.  It is difficult enough for a person to remember
his own past accurately, but it is decidedly more difficult for a person
to recall accurately someone else's past.  Some informants were unable
to keep track of their former classmates as they moved from one job to
another and from one place to another.  So retracing was used as a
device for getting accurate information and for resolving conflicting
pieces of information.

### Confirmation of Data

In our endeavour to get accurate data we tried as much as possible
to confirm the information we had obtained through our tracing.  This
confirming strategy was incorporated in our tracing in the following
ways:

(i)  As a rule, more than two school leavers from each of our sample
schools were interviewed to give fresh information and at the same time
to correct or confirm the previously obtained information.  The big
institutions in and around Nairobi proved very useful in this respect.

(ii)  The Ministry of Education's Sixth Form census was useful as
a source of new information and for confirming some of the results we
had already obtained through tracing.

(iii)  Letters were sent to employers and training institutions all
over the country to obtain confirmation of information we had obtained
from other sources.  In Nairobi, telephone calls to some of the
employers and personal visits to their offices were possible.  We
received a great deal of co-operation from employers, heads of training
institutions and other organizations who had contact with school leavers,
and in only a few cases were our letters not answered.

However, we did encounter some problems arising from school
leavers changing their names after leaving school.  We also found that
the records maintained by Kenya employers rarely indicate where

an employee has come from or where he has gone after leaving this
employment. But despite these difficulties, we were able to obtain
accurate information on most of the school leavers in our sample.
We were unable to trace only about 9 per cent of our sample. (4,pp.49-51)

The high response rate (91%) for such a complicated study is a tribute
to the diligence of the researchers and to the effective combination of different
data collection strategies. Nevertheless, the study does illustrate the type of
analysis problems which are introduced whenever response **rate** is less than 100%.
Kinyanjui reports a table which shows that nearly one out of   five (17%) of
the poorest students remained untraced, compared to only 3% of the superior
students who could not be traced. Thus any comparisons between the poorest
and the superior students is affected by the substantial differences in response
rate. The study reports, for instance, that equal proportions of the very best
and the very poorest students (2% of each group) are in private sector training
programmes, suggesting that these programmes do not provide a greater avenue
into employment for the poorer than for the better students. Which is what one
might have expected. Yet if all the untraced poorer students were to be in
private sector training programmes, perhaps typing schools or apprentice
programmes in the informal sector, then the poorer students would outnumber
the better students by a ratio of nearly 10 to 1. This would suggest a very
different set of relationships between school performance, private sector training
programmes and employment opportunities.

Of course this assumption cannot be made. The difficulty introduced
by non-respondents in any study is that there is no way to know their
characteristics. In the present example, there is no way to know whether the
untraced students are employed or unemployed, in or out of training programmes,
alive or dead. ...nd when the non-respondents come disproportionately from a
particular group in the sample, then analysis suffers.

In commenting on this issue in the context of the Tracer Project we
mean only to illustrate a problem, not to criticise the Tracer Project. As
emphasised earlier, the researchers demonstrated considerable data collection
skill, mixing several strategies in order to find such a high proportion of
students who took the E.A.C.E. as long ago as 1965.

Observation

Initially it might seem that observation is such a simple research tool that no special commentary is called for. This simplicity is deceptive. Observation is a subtle and exceedingly demanding method of collecting social data. Of course this is not so if you are using observation in a casual way, the way any curious person might use observation. But if observation is intended to produce systematic, quantifiable data, then it is far from a simple method.

The trained social scientist **observes** according to established rules. Only in this way can he be assured that the data collected reflect a social condition or social behaviour accurately. The social scientist is in this regard similar to the natural scientist who uses observation. Thus the astronomer does not randomly gaze at stars, as most of us do on a cloudless night. Rather the astronomer observes according to **a conventional set of** scientific rules of observation. Only in this manner can his observations be used to measure and record the characteristics of celestial objects or conditions.

Before turning to the rules of observation in social science we note that there are three major types of observational research. First is what is usually called participant observation. The researcher participates in the events he intends to study. Anthropologists, for instance, sometimes live with the peoples they are studying and as much as possible become participants in the life of the village, perhaps being initiated into an age-set. Urban sociologists studying street gangs have used the same technique, living the life of a gan**g** member for one or two years as a way in which to fully record the social behaviour of such groups.

It takes a very dedicated social scientist to become a participant observer. As much as two or three years might be invested in collecting data. A modified and less demanding form of participant observation is used whenever a political scientist participates in a political demonstration in order to study other participants or a sociologist participates in a meeting of the village cooperative to study local leadership. Such modified particip**ant** observation is not so demanding, but then neither does it generate the wealth of information produced by becoming a fully participating member.

The second major type of observational research is non-participant observation. Here the social scientist observes without participating. A political scientist might study parliamentary debates from the visitor's gallery; he would be observing but not participating. The child psychologist might observe the interaction between mother and child, and of course would not be an actual participant. In a study of teacher methods, the observer might watch teachers conducting their classes. There are many research problems for which the techniques of non-participant observation are **appropriate**; some further illustrations are presented below.

The third major type of observational research occurs in laboratory conditions. In one version a group is brought into a social laboratory, given a task to perform (solve a complex ethical problem, for instance), and then the observer records the social interaction which occurs. Who exercises leadership in the group? What is the group process through which the problem is discussed and resolved? And so forth. In this type of research the observer normally remains hidden, usually behind a window through which he can see but not be seen. This type of experimental research is infrequently used in East Africa, and need not be commented on at length. A less experimental version of laboratory research often occurs with young children. A playroom is kept under observation, again often through a one-way window. Certain kinds of play equipment and toys are left in the room. The observer records the dynamics of play, perhaps by watching which children use what kinds of equipment and how they use them in play situations.

Of the three major types of observational research, non-participant observation is most frequently used in East Africa and is the type on which we shall now concentrate. The rules of observational research, however, are applicable to all three types.

The main rules to be followed are four. First, the observations serve clearly formulated research purposes. Second, the observations are planned systematically. The researcher knows what he intends to observe and, equally important, how he intends to observe it. Third, the observations are systematically recorded. Unless this rule is followed, the data collected from observations cannot be easily quantified, and **we are interested in techniques** which generate quantifiable data. The fourth and most exacting rule is that

the observations should be subjected to controls. This is especially important because it is easy to see what we want to see. And what distinguishes scientific observation from more casual observation is the care taken by the trained observer to check the accuracy of his observations.

A recent student demonstration at the University of Nairobi was described by one newspaper as involving "4000 milling, angry strikers" and by another newspaper as involving "2500 student protesters". This example illustrates the problems of observational research. Two observers (newspaper reporters, presumably) witness the same event and yet see it very differently, depending, one can surmise, on the editorial policy of their respective newspapers. To convert observation into a reliable technique of social research requires a method of checking the observation. The trained social scientist uses such methods, while the casual observer does not.

## Uses of Observational Research

One advantage of observational research over survey or documentary research is that behaviour can be recorded as it occurs. This is particularly valuable if the respondent is likely to distort what happens when he is detached from the event. For instance, in a survey one might ask parents how they handle situations when their children become stubborn or violent in response to parental discipline. Most parents would report that they attempt to reason with the child and use methods which avoid escalating the situation into a parent-child confrontation. These parents probably believe that they have honestly described what they do. But if one observes enough of these situations he learns that parents lose their temper and very often do exactly what they claim to avoid. It is not that the survey data generate dishonesty; it is that detachment from a highly emotional situation often leads to selective recall of the situation.

A great advantage of observational research is that reported behaviour and actual behaviour can be compared. Such comparison might be important in the type of parent-child research just noted. Another illustration is the study of favouritism by persons in authority positions. A survey might ask the Speaker of Parliament whether he favours members of particular tribes during the question period, and he would most certainly reply that his position demands neutrality and fairness. The researcher might then observe the question period and learn that members from particular tribes are systematically favoured.

Related to this point is the use of observation to record behaviour
the subject is unwilling to report.  Store clerks, for example, will not admit
to racism in how they treat customers.  But an observer watching a clerk for
a few hours might well see many instances of racism; perhaps the clerk waits
on Europeans first (or last) or takes more (or less) effort in satisfying their
requests.

Finally, observational research is useful when the subjects cannot
give verbal reports.  Of course this is why observational techniques are so
often used in studies of child development.  It also accounts for the wide
and often effective use of observation in studying animal behaviour.

## Problems of Observational Research

These advantages notwithstanding, one should be aware of three major
difficulties with observation as a research tool.  First, it is costly.  It can
absorb a great deal of research time.  This is especially so when the single
researcher wants to conduct all the observations himself.  And if the observa-
tions are being carried out by a team of researchers, there are other problems.
It is difficult enough to standardise an interview from one researcher to the
next, but it is even more difficult to standardise observations.

The second drawback of observational techniques is that many things
of interest to the researcher are not observable.  This is true of attitudes,
for instance.  There is no way to observe the attitude of a voter toward political
leaders or the attitude of a farmer toward the local extension officer (though one
can sometimes infer attitudes from observed behaviour).  Then there are things
which are too private to be observed.  A study of birth control practices will
necessarily depend on reported rather than observed behaviour.  And there are
things which cannot be observed because they happen over a long period of time.
If the researcher wants to study the careers of civil servants, he will ask them
about the posts they have held.  He will not design a study which depends on
observing them over a period of twenty years.

Finally, we return to a point mentioned earlier.  Observational **data are**
subject to many errors of **misperception** and misrecording.  The observer, even
when trained, can see what he wants to see.  Although this is true also of survey
data and documentary data, the role of the researcher is less pronounced than in
collecting observational data and thus the room for error is less.  In using
observational techniques the researcher is particularly subject to self-deception.

CHAPTER 6

QUANTIFICATION AND LEVELS OF MEASUREMENT

Social scientists use numbers in two different ways. The simpler application is in descriptive statements. Instead of saying that "Most senior civil servants are well-educated", the investigator may prefer more precision and thus report that "Of the senior civil servants, sixty per cent have university degrees, another thirty per cent completed H.S.C., and the remaining ten per cent finished secondary school". Descriptive statistics such as frequency counts, averages and proportions, along with auxilliary devices such as tables, charts and graphs, replace terms such as 'some, most, a few, nearly all, several'. The early stages of quantification in a particular field of inquiry involve an increase of descriptive statistics.

Of course common sense should be used. Not every statement in every report need be made precise in this way. But when the author is basing an interpretation on actual counts or proportions, he owes to the reader as precise a statement of these quantities as is available.

We know that numbers can give a false sense of precision. They are subject to errors of collection, interpretation and reporting. But in this regard quantification does not differ from more vague and imprecise descriptions. Because nearly all social science implies assigning quantities to what is observed or recorded, to use terms such as most or few is as subject to error as to use actual counts and proportions. Indeed, quantification alerts the researcher to the many sources of error and might produce more careful work than that produced when numbers are avoided.

The use of quantification in descriptive statements has a further advantage. Authors sometimes describe a characteristic of their population as if it were universally shared: "Bureaucrats in Kenya are notable for their indifference to complaints by Asian citizens." This is a very misleading statement for the reader has no way of knowing whether the author intends to describe all bureaucrats or only some proportion. Were the author forced to think in terms of descriptive quantification he would not be allowed the luxury of a statement which can be interpreted in more than one way.

The second application of numbers in social science research involves relationships and associations between variables. In addition to frequency counts and proportions, we use associational or correlational statistics. This

involves ranking or otherwise differentiating the thing being studied in
terms relevant to the research, as when it is said that regions differ in
the number of secondary schools they have, extension services differ in how
productive they are, marketing boards differ in their rate of return on
investments and so on. When two or more variables have been ranked, the
direction and strength of association between them can be determined.

Consider the hypothesis, "Development plans prepared by local staffs
have a better chance of being implemented than plans prepared by expatriate
advisors". The independent and dependent variables in this hypothesis could
be quantified in terms of descriptive statistics; thus implementation is measured
by a frequency count of the programmes put into effect within a given time
period, and preparation is measured by a ratio of local staff to expatriate.
If a number of development plans were ranked from low to high implementation
and from low to high local staff preparation, the association between the two
rankings would provide the test of the hypothesis.

Once a field of inquiry has sharpened its tools for descriptive
statistics, then it is possible to test for relationships between variables.
This of course is the great value of quantification. This testing for relation-
ships allows the researcher to locate a condition which is causing a break-down
in a marketing procedure, or allows him to learn why two seemingly similar
training programmes are having greatly different effects on the morale and
subsequent performance of the trainees. Quantification is not a remedy for
the problems facing the administrator in a developing nation, but it often can
reveal those problems in a manner which makes solution possible.

The chapters which follow review materials relevant to both descriptive
statistics and statistics used in testing relationships among variables. As
background to these materials it is useful to have an understanding of what
methodologists call levels of measurement. This notion in turn rests upon an
understanding of variables and categories.

## Variables and Categories

The term variable is widely used in social science. Fortunately it
has a common-sense meaning. A variable is a characteristic which varies across
a defined set of units. A defined set of units is most often a population of

individuals (individuals then being the units), as in students at the University of Dar es Salaam, Ugandan farmers, civil servants in the Ministry of Agriculture, tourists and so forth. Not all defined **sets of units need be persons.** Here are other examples:

| Defined Set of Units | The Unit |
|---|---|
| Members of the United Nations | Individual nations |
| Provinces of Uganda | Specific provinces |
| Automobiles on Kenyatta **Ave.** | Each automobile |
| Planets of the sun | Each planet |

Technically, the meaning of varies is that the values of the characteristic differ among the units in the set. Thus we are in the presence of a variable whenever units differ in terms of some characteristic. Thus:

| Defined Set of Units | Characteristic | **Diffe**ring Values (or scores) |
|---|---|---|
| Univ. of Dar es Salaam students | Sex | Male/female |
| Ugandan farmers | Land-ownership | Not much land to a lot |
| Civil servants | Location | In/out of Kampala |
| Tourists | Nationality | German/American/etc. |
| Members of U.N. | Size | Small to big |
| Provinces | Social services | Few to many |
| Automobiles | Speed | Slow to fast |
| Planets | Distance | Near to far |

A common-sense understanding of the term variable leads naturally enough to a common-sense understanding of the term category. Categories are those things (scores or value classifications) which describe just how the units vary. In the above examples, the categories derive from the things listed under differing values. The categories for the variable sex would be male/female; for the variable land-ownership would be $\frac{1}{2}$ acre, 1 acre, 2 acres, etc.

Three points should be kept in mind about categories:

a) They must sort the population (the set of units) into mutually exclusive groups. One cannot have categories which allow some members of the population to belong to more than one category.

b) The categories must provide an exhaustive classification. Some units cannot be left over when the classification has been completed. (Sometimes this rule is obeyed simply by creating the category of 'others' or 'miscellaneous' which spoils **research** if too many units fall into such a catch-all. But there are certainly times when it is appropriate. The **different nationality categories used to classify** tourists might list the major groups, and then have a 'miscellaneous' to pick up the hitchhiker from Iceland, etc.

c) The number of categories a variable can have is
   often the choice of the researcher. Age, for
   instance, could be two categories: persons 30
   years and younger and persons over 30; or it could
   be a dozen categories: persons from 1 to 8, from
   9 to 16, from 17 to 24, etc.; it could be several
   hundred categories: persons from one month to three
   months, from three months to six months, etc.
   Not every variable is so flexible. Sex lends
   itself to only two categories. There are only
   seven provinces of Kenya, and thus classifying by
   province is limited to seven categories.

The actual number of categories into which a variable
is divided in research will depend on a good many
factors, not least of which is the subtlety of the
measuring instrument. The beginning researcher is well-
advised to keep the number of his categories manageable.

## Levels of Measurement

Now that we have discussed the meaning of a variable and its cate-
gories, we consider how variables differ in their level of measurement.
There are four levels of measurement; and it is best simply to describe them
before trying to explain why we make these distinctions.

The four levels of measurement (some books call them scales) are
ratio, interval, ordinal and nominal.

Ratio Scales:  A variable which can be measured as a ratio scale is age.
Anyang is 20 years old; his younger brother, Apindi, is 10 years old.  Once
we have classified them in terms of age, four conclusions can be drawn.
First, we can say that they have different ages.  Second, and more useful,
we can say that Anyang is older than Apindi.  Third, we can say how much older
Anyang is than Apindi, as we say that he is ten years older than Apindi.
Fourth, we can say by what ratio Anyang is older than Apindi.  He is twice
as old, which is another way of saying that the ratio of their ages is
2 to 1.  It is this last conclusion which leads us to say that age is a ratio
measurement.

Although it is important to know what a ratio scale is, true ratio
scales are not often found in social research.  For a variable to qualify as
a ratio scale it must have a non-arbitrary zero-point.  Of course age meets
this qualification.  You can be zero years old, if only for a fleeting second.
The physical sciences are better equipped when it comes to ratio measures
(which is why purists claim that only the physical sciences are true sciences).

Weight, mass and velocity can be measured as ratio scales. It makes sense
to say that 100 kilos is four times as heavy as 25 kilos; that there is such
a thing as non-mass (the non-arbitrary zero point); or that the ratio of speed
between a jet and non-jet is 3 to 1.

But before claiming too little for the social sciences, we should add
that very often our measures can be treated as if they were ratio scales.
That is, we act as if we have a non-arbitrary zero point on a measure of
farmer innovativeness, departmental efficiency, student radicalism or urban
congestion.

A ratio scale is the highest level of measurement, and all properties
belonging to the remaining three scales belong to the ratio scale as well.

Interval Scale: This level of measurement allows the first three conclusions
associated with the ratio scale, but not the fourth conclusion. The most
frequently cited illustration of an interval scale is temperature. If Anyang
has a temperature of 101.0 degrees F, and Apindi's temperature is a normal
98.6$^{\circ}$, we know that the brothers differ in their temperature, that Anyang has
a higher temperature, and that Anyang's temperature exceeds that of Apindi by
2.4 degrees. But we cannot compute a ratio, for the simple reason that a
Farenheit scale does not have a true zero point, but only an arbitrary one.
(You can easily prove this to yourself when you realize that zero on the
centigrade temperature scale is 32 on the farenheit scale.

The term interval derives from the third conclusion suggested above.
That is, you can establish the interval between any two points on the scale.

Ordinal Scale: For two reasons it is the ordinal scale, or ordinal level of
measurement, on which we want to concentrate. One reason is that most variables
in the social sciences are ordinal measures; the second is that a discussion
of how variables order the units measured is basic to comprehending what social
scientists do.

An ordinal measure meets the two easier criteria of measurement, but not
the harder two. An ordinal measure tells us that the units have different
scores and that one unit has a higher score than the other. It cannot tell us

anything about the size of the interval between scores or about the ratio by which one unit is higher than another.

Social status is a good example of an ordinal scale.  We say that Okelo, a member of parliament, has a higher social status than Kuria, a bank clerk.  The measure of social status (occupational prestige) tells us that **Okelo and Kuria belong to different categories, and** that one category is higher than the other.  It does not tell us that the status interval between Okelo and Kuria is so many points, and it does not tell us that Okelo has twice the status or three times the status of Kuria.

An ordinal scale is a system of grading or ranking.  Our vocabulary reflects this: we talk about one thing being more than, higher than, greater than or harder than another thing.  A great many social science concepts are ordinal (implicitly):

> Stability ("One form of government is more stable than another.")
>
> Development ("A region with a network of roads is more developed than one without.")
>
> Violence ("There is more violence in personal relations now **than** there was 20 years ago.")
>
> Competition ("A multiple-party system encourages more competition than does a single-party system.")
>
> Power ("The executive is more powerful than the legislature.")
>
> Innovation ("Wealthier farmers tend to be more innovative than poorer farmers.")

For each of these broad concepts -- stability, development, violence, competition, power, innovation -- we could devise measures which would rank-order governments, regions, persons or whatever from low to high.  It is the rank-order idea which is basic to an ordinal scale.  (It is also basic to an interval and ratio scale, but as we have seen, interval and ratio scales do more than just rank-order.)

Some points to remember about ordinal scales are:

1.  An ordinal scales is asymmetric.  Asymmet**ry** simply means that if X is greater than Y, then Y cannot be greater than X.  If Okelo has higher status than Kuria then Kuria cannot have a higher status than Okelo.

At this point, we should list the notation used to make these
kinds of statements:

> Greater than

< Less than

$\geq$ Greater than, or equal to

$\leq$ Less than, or equal to

$=$ Equal

$\neq$ Not equal

$\ngtr$ Not greater than

$\nless$ Not less than

Using this notation, the idea of asymmetry is written as follows:

If X > Y, then Y $\ngtr$ X.

2. An ordinal scale is transitive. If Okelo has higher status than Kuria,
and if Kuria has higher status than Mbigi, then Okelo has higher status than
Mbigi. Or, using the notation,

If X > Y, and if Y > Z, then (necessarily) X > Z.

Or,　　　　If X < Y, and if Y < Z, then (necessarily) X < Z (or Z > X).

3. An ordinal measure presumes that the variable measured has an underlying
continuous distribution, even though that distribution is not actually
measured. This point can be shown through example.

In a study of agency effectiveness five separate agencies are
investigated. Effectiveness is a variable with, in theory, categories
stretching from totally ineffective to completely effective.

|ıııııııııııııııııııııııııııııııııııııııııııııııııııııııııııııııııııııııı|

Totally　　　　　　　　　　　　　　　　　　　　　　Completely
Ineffective　　　　　　　　　　　　　　　　　　　　Effective

Each mark represents a category of effectiveness. If we could actually
measure effectiveness so precisely, we would have a ratio scale. (There is
a non-arbitrary zero-point, total ineffectiveness, and the interval between
each category mark is known.) Let us assume that our five agencies are located
as follows:

| Agency | Agency | | Agency | Agency | | Agency |
| A | B | | C | D | | E |
| ↓ | ↓ | | ↓ | ↓ | | ↓ |

If we knew them to be so located we indeed would have a ratio scale (or at least an interval scale) because we can see the distance between any two agencies and could even compute a ratio of effectiveness.

Unfortunately our measures are normally very crude, certainly so when studying a concept so elusive as effectiveness. What our study would actually report is that the five agencies can be rank-ordered from low to high in terms of effectiveness:

<div align="center">

Least Effective -- Agency A

Agency B

Agency C

Agency D

Most Effective  -- Agency E

</div>

This rank-order does not tell us anything about how much more effective one agency is than another, that is, it tells us nothing about the distance (or interval) separating them. The rank-order is compatible with dozens of different possible distributions, of which three are illustrated here:

```
         A                                   B   C   D   E
1)  _____

              A  B  C  D  E
2)  _____

       A  B          C          D          E
3)  _____
```

In the first distribution, there is one very inefficient agency and four quite efficient ones. In the second distribution, the agencies are practically the same in effectiveness, and all about average. In the third distribution, the agencies are spread out, but all toward the lower half of the effectiveness measure.

The point is, an ordinal measure does not tell us which of these distributions is the actual one. It only tells us that A is less effective than B, B less than C, and so forth.

Social scientists normally assign scores to the different units.
Perhaps Agency A gets a score of 1 (indicating least effective), Agency B a
score of 2, and so forth to Agency E with a score of 5. These scores indicate
rank-order and that is all. They do not have common arithmetic properties.
That is, one cannot say that the combined effectiveness of Agency A and B (a
score of 1 plus a score of 2) is equal to the effectiveness of Agency C (with
its score of 3).

The best way to understand this is by realising that one can change
the scores at will as long as the same rank-order is kept, and nothing has
changed in the measurement. Give Agency A a score of 1, Agency B a score of
5, Agency C a score of 10, Agency D a score of 15 and Agency E a score of 20.
Other than telling us the low to high ranking of the agencies, the scores mean
nothing.

4. An ordinal scale not only obscures the underlying continuum, it treats
clusters of units as if they were ties. Here is a question used in a survey
of 2,000 Kenyan secondary students.

> Do you agree or disagree with the idea that only persons who
> are at least literate should vote in national elections?

Strongly Agree _____

Agree _____

Disagree _____

Strongly Disagree _____

The agree-disagree options establish an ordinal scale, for they permit ranking
the students in terms of levels of agreement. But note that only four options
are given. This means that there will be a great many ties. In fact, of the
2,000 students, 140 strongly agreed, 297 agreed, 403 disagreed and the remaining
1,160 strongly disagreed. Each category has a larger number of ties within it.

But these ties are the result of a non-sensitive measuring instrument.
It is not likely that, for instance, the 1,160 students who ticked 'strongly
disagree' feel exactly the same. Within that group are probably some who feel
so strongly that were literacy tests used to determine voter eligibility they
would join an active opposition political movement. Others within this  group

disagree simply because that seems the appropriate kind of response to make
in a society which calls itself democratic, but they do not actually feel
all that strongly.  There probably are others between these extremes.

Whatever the true distribution, the research instrument treats
all 1,160 who tick 'strongly disagree' as if they felt identically.  Any
measure which has only a few categories must necessarily presume ties across
the units being measured.  A study which classifies regions as low rainfall
and high rainfall is treating all regions within either of the categories as
if they had the same amount of rainfall, as if they were tied.  But this is
unlikely.  There are many degrees of low rainfall just as there are many
degrees of high rainfall.  Nevertheless, ordinal measures which usually have
a limited number of categories do end up treating cluster of units as if they
are ties.  The more ties produced by a measurement, the less sensitive the
measurement.

Nominal Scale:  The last and least powerful type of measure is the nominal
scale or, more appropriately named, the nominal classification.  A characteristic
which is measured at the nominal level has categories which classify but do
not rank the units.  Sex is a nominal measure.  We observe that there are two
categories, male and female, but do not presume that one of the categories is
somehow more than, or higher than or greater than the other.  (If we say that
males are generally stronger than females, the variable being measured has
been shifted from sex to strength.  And strength can of course be an ordinal
measure.)

Other examples of nominal measures include region, race, tribe and
nationality.  For each of these variables, we classify persons into different
groups but we do not assume the classification somehow ranks one group over
another.  The distinction between nominal measurement (so-called because it names
different categories) and the other three levels of measurement is the distinction
between differences of kind and differences of rank.

The distinction between differences of kind and differences of rank
will become clearer if we consider the implications of blurring the distinction.
Some of the most unjust political regimes in history have been based on blurring
just this distinction.  In South Africa today, for instance, a difference of kind

(race) is transformed by the political ideology, and the power which backs it up, into a difference of rank. Whites are not said to be different from blacks but are said to be superior. The classification by race is given an artificial rank-order. This is what leads most of us to conclude that racist ideologies (or sexist ideologies or tribalist ideologies) are unjust. They take perfectly normal social differences (white/black; male/female; Kikuyu/Luo) and attempt to convert them into differences of rank rather than kind.

## Summary

We have now presented four levels of measurement, differentiated by their properties. All four scales have the property of classification -- they put units into different categories. Three of the scales have the property of ordering -- they rank the units. Only two of the scales have the property of intervals or distances -- they tell us how far apart the units are. And only one scale allows us to compute ratios; this scale has in addition to the other properties the property of a nonarbitrary zero point.

It is useful to have this in table form.

### Four Scales

| Property: | Nominal | Ordinal | Interval | Ratio |
|---|---|---|---|---|
| Classification | yes | yes | yes | yes |
| Ordering | no | yes | yes | yes |
| Distances | no | no | **yes** | yes |
| Nonarbitrary zero | no | no | no | yes |

One reason for bringing these different levels of measurement to the attention of the reader is that social scientists frequently make use of the vocabulary introduced in this chapter. It is common to read in a research report that 'the measures relied on are **ordinal**' or 'statistical procedures used in this analysis presume interval level measurement'. Such statements will make sense to the beginning student of social science methodology or to the consumer of research reports only after understanding measurement levels.

There is a very practical reason for social scientists to distinguish among the different levels of measurement: different statistics have been devised for use in conjunction with different levels of measurement. In the

next chapter two statistics of association are introduced: the gamma measure of association and the Spearman rank-order statistic. Both of these statistics are appropriate when the data are measured at the ordinal level. Other statistics, introduced in subsequent chapters, are appropriate when data are measured at the interval or ratio level.

CHAPTER 7

MEASURES OF ASSOCIATION

Statistics are useful when the researcher wishes to describe or summarise a characteristic of his sample. Often only one characteristic at a time is described or summarised, as in these examples:

> The average income of university graduates in Tanzania is Shs 2,000 a month.
>
> Sixty-four percent of Kenyan children aged between 7 and 11 attend primary school.
>
> The highest civil service salary recommended by the Ndegwa Commission Report is approximately 38 times greater than the salary recommended for the lowest job category in the civil service.

Statistics such as averages, proportions and ratios are effective means of simplifying large amounts of data. Thus a statement specifying the average income of university graduates in Tanzania is much more efficient than a statement listing each individual's income in a sample of 500 university graduates.

Social science researchers, however, are seldom content simply to describe the characteristics of their sample. Most social scientists plan their studies so that they can discover relationships within their sample. It is in the discovery of relationships that the social scientist begins to formulate explanations for why things are as they are. It would be more interesting to learn the relationship between income and education in Tanzania, taking into account all educational levels, than only to know the average income of university graduates. The figure Shs 2,000 a month means one thing if the average income of secondary school graduates is Shs 1,000 a month, and quite a different thing if the average income of secondary school graduates is Shs 1,900 a month. Our understanding of social classes, or classlessness, in Tanzania is furthered by knowing the pattern of relationship between the variables income and education.

Statistics are aids in discovering relationships between variables and in efficiently summarising these relationships. Statistical relationships have two main properties: direction and strength. The direction of a relationship is determined by whether two variables are positively or negatively related. Income and education **are positively** related when persons with higher educations tend also to have higher incomes and persons with less education tend to have lower incomes. An example of a negative relationship might be that between size

of family and proportion of children in the family attending school: if as
size of family increases the proportion of children being educated decreases,
the relationship would be negative.

The strength of a relationship is determined by the frequency with
which the variables covary across the sample.  If all university graduates
have higher incomes than all secondary school graduates, and if all secondary
school graduates in turn have higher incomes than all primary school leavers,
and if all primary school leavers in turn have higher incomes than all un-
educated persons, then the relationship between education and income will be
very strong indeed.  But if, as would probably be the case, some persons with
less education actually have higher incomes than some persons with more
education, then the relationship will be weaker.  The measures of education and
of income covary only up to a point.  And of course if less educated persons
are just as likely to have high incomes as well educated persons there will
be no relationship between the variables.  The one variable will not covary
with the other.

One of the best ways to introduce the concept of statistical **relation-**
ships, and to see what is meant by the direction and strength of relationsh**i̟s,**
**is** to review a rank-order statistic.  A rank-order statistic requires that the
things being studied are ranked in two ordered series.  For instance, a study **of**
education and income would rank every person in the sample from high to low in
terms of education and from high to low in terms of income.  These two ranks
would be compared to see whether the persons at the top of one ranking tend also
to be the persons at the top of the other ranking.

## Spearman Rank-Order Statistic

The example we use compares the ranking of the provinces of Kenya on
two measures: the size of the population and the per cent of university graduates
who come from each province.  In determining the relationship between these two
rankings, we will be learning whether the provinces contribute university
graduates in proportion to the size of their populations.  A perfectly positive
relationship indicates proportional contribution; the absence of a positive
relationship indicates that factors other than population size affect the number
of students each province sends to the university.

The relevant statistic is known as the Spearman rank-order statistic, often symbolised as:

$$r_s$$

It is also sometimes called rho. The formula for the rank-order statistic is:

$$r_s = 1 - \frac{6 \; \Sigma d^2}{N \, (N^2 - 1)}$$

This statistic can take a value from -1.00 through 0 to +1.00, the minus one indicating a perfect negative relationship between the two rankings and the plus one indicating a perfect positive relationship between the two rankings. To take our example, if the smallest province sent the most number of students to the university, the next smallest province sent the next greatest number of students and so forth, the Spearman rank-order would be -1.00. This would be complete disagreement between the two ranks.

If, however, the largest province also sent the greatest number of students, the next largest province the next greatest number of students and so forth, the relationship would be +1.00, or perfect agreement. (In practice, a positive relationship is normally not reported with the sign; thus the relationship would be reported simply as 1.00.)

In computing the statistic the initial step is to rank the provinces in terms of the two measures, as is shown below. A quick inspection of the rankings indicates something less than perfect agreement. Rift Valley, for instance, is the most populous province but only the seventh in terms of university graduates, while the province which has the greatest number of graduates is fourth in terms of size.

| Province | Rank-ordered by population (1969) | Rank-ordered by % of University grad. - 1964-70 | difference | squared difference |
|---|---|---|---|---|
| Rift Valley | 1 | 7 | -6 | 36 |
| Nyanza | 2 | 2 | 0 | 0 |
| Eastern | 3 | 6 | -3 | 9 |
| Central | 4 | 1 | 3 | 9 |
| Western | 5 | 5 | 0 | 0 |
| Coast | 6 | 4 | 2 | 4 |
| Nairobi | 7 | 3 | 4 | 16 |
| North-Eastern | 8 | 8 | 0 | 0 |
| | | | sum $d^2$ | 74 |

(Note that in the "difference column" the total of the negative numbers equals the total of the positive numbers. This will always be the case, and is useful to keep in mind as a check on the arithmetic.)

Visual inspection of the two rank-orders, however, is difficult.
One cannot easily summarise the relationship without systematically comparing
the two ranks. The Spearman statistic provides a method for systematic comparison.
The table indicates the computational steps which first call  for finding and
then squaring the difference for each unit between its score on one rank and its
score on the other rank. The difference for the Rift Valley is -6, because it
is first in population but seventh in university graduates. As always, when a
negative difference is squared the sign becomes positive.

The formula says that the sum of the squared differences is to be
multiplied by six, thus 6 x 74 = 444.

This number should be divided by N squared minus 1 (8 x 8 = 64 - 1 = 63)
times N (63 x 8 = 504). Thus, 444 divided by 504 equals .87. According to the
formula we subtract this from 1 (very important) : 1 - .87 = .13.

For the table presented, the Spearman statistic is .13. There is a
positive but weak relationship between the size of a province's population and
the proportion of university graduates who come from that province. This suggests
to the researcher that factors other than population size affect the flow of
students into the university.

By rearranging the ranks and recomputing the Spearman rank-order
statistic the reader can easily demonstrate that perfect **agreement** between the
ranks would produce 1.00 as the summary statistic. Perfect disagreement between
the ranks would produce -1.00 as the **statistic.** It might be useful to engage in
this exercise as a way of further understanding how a simple summary statistic
such as the Spearman statistic can effectively communicate the direction and
strength of the relationship between two variables.

Rank-Ordering

This brief review of the Spearman rank-order statistic is most
important for the general concept which it implies. It has been stressed in
this text that much of social science involves the search for relationships.
The most common type  of relationship investigated is that between two variables
which characterise the same units of analysis.

This abstract language takes on concrete meaning in the example just presented. The units of analysis are the provinces of Kenya. Each province, or unit, has a different number of people living within its boundaries. Thus the measure population size is a variable, because the units can be shown to vary in the size of their populations. Each province is also measured on a second variable: the proportion of university graduates coming from that province. And again there is variability, because provinces contribute un- equally to the university.

At issue is whether the two variables covary. Is a province which is high on one measure (population size) also high on the second measure (number of students sent to the university)? When the variables do covary then of course we speak of a relationship. If there is no covariability--where a province ranks on one variable bears no relationship to where it ranks on a second variable--then we speak of the absence of a relationship.

Note that the statistic used to determine the direction and the strength of the relationship requires only that the provinces be rank-ordered. It is for this reason that the Spearman is known as an ordinal statistic, so-called because the level of measurement necessary before computing a Spearman statistic is an ordinal scale. As noted in the previous chapter, an ordinal scale classifies the units and then rank-orders them. It does not show the interval between the units.

The proportion of university graduates from the different provinces could be as shown in A, B or C below, and the Spearman statistic would be the same in each case. This is because the order of the units does not change.

| Province | A-Ranking (nearly equal) | B-Ranking (very unequal) | C-Ranking (bunched toward extremes) |
|---|---|---|---|
| Rift Valley | 10% | 2% | 3% |
| Nyanza | 15 | 22 | 21 |
| Eastern | 11 | 5 | 5 |
| Central | 16 | 35 | 25 |
| Western | 12 | 7 | 7 |
| Coast | 13 | 12 | 19 |
| Nairobi | 14 | 17 | 20 |
| North-Eastern | 9 | 0 | 0 |
| | 100% | 100% | 100% |

What these hypothetical rankings show us is that in ordinal statistics it is the ranking which matters, not the absolute values.  This point has far-reaching consequences for the type of data collection strategies chosen by a researcher.  If the researcher intends to use only ordinal statistics his data collection need not measure precisely as long as he correctly ranks the things being studied.

The distinction between measurement and ranking (a distinction which should not be pressed too far) is often not appreciated  even by social science researchers, but it has important practical ramifications for researchers in East Africa.  Because precise measurement is often difficult under the research conditions frequently encountered in East Africa, especially in rural East Africa, the social scientist might be well-advised to use ordinal statistics. His **task** then becomes the correct ranking of his units.

Here is a concrete application of this principle.  A researcher in Tanzania intends to investigate morale in Ujamaa villages to see whether there is a systematic relationship between morale and productivity.  His measure of productivity is yield per acre cultivated, and we will presume that the measure correctly ranks the villages in his sample.

Measuring morale is of course a good bit more complicated.  The researcher has two choices.  He can visit his sample villages, talk informally with the local leadership, observe the villagers as they go about their tasks and then assign to each village a 'morale rank' based on his impressionistic evidence.

He could also design a more systematic study which includes attitude surveys of peasants and leaders in each sample **village**, counts of the number of peasants who attend community meetings and contribute to community projects, data on the rate of in and out migration for each village and similar indicators of morale.

There are a number of important points to make about this hypothetical research situation, points which will help put the issue of social science measurement into a broad perspective.

1.  The casual data collection is much less expensive than the more systematic data collection, and cost is always a factor in designing a study.

2.  There is a 'true' ranking of the villages according to level of morale, but neither data collection strategy is certain to discover the true ranking.  The transition from a concept such as morale to specific research operations always presents difficulties and introduces ambiguities.  We normally assume that more extensive and systematic data collection increases the chances of finding the 'true' ranking, but we can never be certain.

3.  If the researcher wishes only to use ordinal statistics and _if_ the casual collection of evidence provides the same ranking of villages as the more systematic collection, then the casual methodology would be just as good as the more demanding methodology.  For ordinal statistics, the criteria of good measurement is not precision but the correct ordering of the units studied. A less expensive but adequate methodology is normally preferred to a more expensive and equally adequate methodology.

4.  However, the more precise and systematic measurement strategies will probably allow the application of more advanced statistics (as discussed in the next three chapters), and thus lead to more extended analysis.

5.  The greatest danger of the casual data collection is that it is not easily replicated.  There is no way a second researcher, who might dispute the findings, can visit the same villages, apply the same research  techniques and thus check the ranking of the first researcher.  By their very nature, informal and non-systematic data collection methods cannot be sufficiently described to the research community so that members of the community can evaluate the methods.

In contrast, the systematic study reports a methodology which can be replicated.  Another researcher can re-visit the same sample, or take a different sample of villages, and apply the same survey and measurement instruments to see if similar results are obtained.

In the long run, a methodology which can be replicated is strongly to be preferred over a methodology which cannot be replicated.  It is through replication that confidence in findings can be increased.

The purpose of understanding a rank-order statistic, such as the Spearman statistic, is that it shows **what** is meant by a relationship in social science. The reader can see that a sample of units is being ranked on two separate variables; and that the statistic is measuring the degree to which the two variables rank the **units similarly.** Although more advanced correlational statistics reveal less clearly this basic relationship, they are based on very similar principles. Does the score on one variable increase as the score on a second variable increases? If so, there will be a positive relationship between the variables. Or, does the score on one variable decrease as the score on a second variable **increase? If so, there will be a negative relationship. Or, does** the **score on one variable bear no systematic relationship** to a score on a second variable? If so,a measure of association or of correlation will show zero (or near zero), indica-**ting no relationship.**

Kendall Coefficient of Concordance

The Spearman rank-order statistic measures the direction and strength of association between two rankings. But sometimes in social research it is important to learn whether there is association across more than two rankings. One way to do this would be to compute the correlation between each pair, and then compute the average of all pairs. But this could be very tedious if very many rankings were involved. A simpler method is provided by the Kendall Coefficient of Concordance, often symbolised as W.

This statistic takes values between 0 and +1.00. That is, if there is perfect agreement among the rankings, the association will be +1.00. If there is no agreement at all, the association will be 0. This statistic cannot give a negative correlation, because when more than two sets of rankings are involved they cannot all disagree completely. (You can easily prove this to yourself by trying to rank three different variables in a way which would show perfect dis-agreement among them.)

The formula is:

$$W = \frac{s}{(\frac{1}{12} k^2) \times (N^3 - N)}$$

s = sum of the squares of the observed deviations from the mean
k = number of sets of rankings
N = number of things being ranked

Computation: **Imagine a study to** determine whether three agricultural experts agree among themselves as to the crop most useful for a particular soil and locality. Each expert is presented a list of six crops and asked to rank

them from one to six in terms of suitability. What we wish to know, then,
is how much agreement there is among the experts. In this example, there
are three sets of rankings (k) and six things being ranked (N). The data
are set up as follows:

| Crop: | Three Rankings A | B | C | Total Score Each Crop | Mean Score | Diff. | Difference squared |
|---|---|---|---|---|---|---|---|
| Coffee | 1 | 1 | 6 | 8 | 10.5 | -2.5 | 6.25 |
| Tea | 6 | 5 | 3 | 14 | 10.5 | 3.5 | 12.25 |
| Maize | 3 | 6 | 2 | 11 | 10.5 | .5 | .25 |
| Pyrethrum | 2 | 4 | 5 | 11 | 10.5 | .5 | .25 |
| Sisal | 5 | 2 | 4 | 11 | 10.5 | .5 | .25 |
| Forest | 4 | 3 | 1 | 8 | 10.5 | -2.5 | 6.25 |
| | | | | $63 \div 6 = 10.5$ | | | $25.50 = s$ |

Once we have s (sum of squared differences of each score from the mean), we
simply put it into the formula:

$$W = \frac{25.5}{(\frac{1}{12} \times 3^2) \times (6^3-6)} = \frac{25.5}{(8.33 \times 9) \times (216-6)} = \frac{25.5}{(.75) \times (210)}$$

or $\frac{25.5}{157.5} = .16$

The coefficient of concordance is .16, indicating that there is not much agreement
among the three experts regarding the most suitable crop.

An application of Kendall's coefficient of concordance appears in a
research report based on survey data collected from a national sample of secondary
school students in Kenya. In this particular research report the authors were
investigating patterns of regional heterogeneity and homogeneity. One research
hypothesis was that secondary students would exhibit regional diversity in
certain political values, but would exhibit a great deal of similarity in their
perceptions of Kenya's occupational structure. To examine the second part of the
hypothesis the students were asked to indicate their own job preferences as well
as to rank jobs in terms of their importance to the development of Kenya. The
table presented below is based on the question, "Which do you think are the three
most important jobs to help the development of Kenya?" The table demonstrates
that secondary students in six different regions perceive the relative importance
of jobs in highly similar ways. The Kendall coefficient is .94, indicating nearly
total agreement in how the jobs are ranked by students of the six areas. A high
coefficient (.92) is also found when students are asked to indicate their job
preferences.

**Jobs** Rank-Ordered according to their Perceived Importance in the
Development of Kenya by Province

| Province | Teacher | Farmer | Scientist | Nurse | Politi- cian | Police- man | Factory Owner | Business- man | District Commissioner |
|----------|---------|--------|-----------|-------|--------------|-------------|---------------|---------------|-----------------------|
| Western | 1 | 2 | 3 | 5 | 4 | 7 | 6 | 8 | 9 |
| Nyanza | 2 | 1 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Rift | 1 | 2 | 3 | 4 | 6 | 5 | 7 | 8 | 8 |
| Central | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| East | 1 | 1 | 3 | 2 | 5 | 6 | 7 | 9 | 8 |
| Coast | 1 | 2 | 3 | 4 | 8 | 6 | 7 | 5 | 9 |

Jobs rank-ordered on basis of proportion selecting it as the most
important. Kendall's Coefficient of Concordance: .94.

The utility of this statistic is that it efficiently summarises a large
amount of data. The reader need not examine the whole array of data in detail.
All he need learn is the strength of the coefficient which measures agreement in
the rankings. If in the present case the coefficient approached zero the authors
would conclude that regional differences in values resist the homogenising effect
of a national economic system. Instead, the coefficients approach 1.00, leading
the authors to conclude that "Kenyan secondary school students, whatever their
regional origin, traditions or politics, are agreed on the types of jobs which
they would like to have on leaving school and in the kinds of jobs which they
relate to the national development of Kenya." (5, p.112)

Table Statistics

The most common form of data presentation in the social sciences is
tables which present proportions. Normally these tables tell us what proportion
of a sample has a particular characteristic. Such a table is illustrated here:

Occupational Background of Candidates Selected
By the National Executive Committee (TANU) --
Tanzania Election, 1970

| | Percentage |
|---|---|
| Farmers | 26 |
| Other government employees | 13 |
| Teachers | 12 |
| Incumbent MPs | 29 |
| Employees in TANU and affiliates | 10 |
| Others | 10 |
| | 100 |

n = 202

This table shows at a glance the occupational background of the candidates in the Tanzania election of 1970. Reading a table such as this presents no difficulties, at least not if the table is correctly constructed. This table has a heading which tells the reader exactly what data are being presented. It also shows the number (n = 202) on which the percentages are calculated, in the event that some readers may wish to know the absolute numbers in each category. For example, we can determine that the candidates include 53 persons who are farmers (.26 x 202 = 53), and 58 persons who are incumbent members of parliament (.29 x 202 = 58).

The occupational background table is a single-dimension table. It shows how the sample is distributed on only one variable. If such a table is adequately labeled, and if it is clear in what direction the table is percentaged, it is easily read.

Difficulties in table reading occur when the researchers are presenting several different pieces of information simultaneously. It then becomes necessary for the reader to proceed slowly, determining just what the researcher is trying to show. Here, for example, is a more complicated table, though again one which uses only proportions.

Political Attitudes Expressed by Kenyan Secondary
School Students: Rank Ordered by Province

| Province | n | Money should not go from Developed to Undeveloped Regions | Sees Government Doing Many Things for Family | Not Good When People Criticise Government |
|---|---|---|---|---|
| Central | (303) | 63% | 31% | 34% |
| Rift Valley | (150) | 51% | 25% | 33% |
| Eastern | (212) | 44% | 16% | 26% |
| Coast | (112) | 43% | 20% | 29% |
| Western | (128) | 30% | 15% | 14% |
| Nyanza | (224) | 43% | 14% | 14% |

To read this table it is necessary to separate its various components.

1. The heading tells the reader that data on political attitudes of Kenyan secondary school students is being presented.

2. The far left column tells the reader that the sample has been classified by province, and thus the reader will be able to compare the political attitudes of students across the different provinces.

3. There is also a column labeled (n), which tells the reader how
   many cases form the basis on which the percentages are calculated.

4. The next three column headings indicate the political attitudes
   (a footnote gives the exact wording of the survey questions).

5. Having examined the basic framework of the table, the reader would
   turn attention to the data themselves. For example, we see that 63
   per cent of the Central Province students, compared to only 30 per
   cent of the Western Province students, believe that money should
   remain in the developed regions of the country.

6. Note that this table shows no rows or columns which total 100 per
   cent. In such a table each entry stands on its own. Thus to read
   that 63 per cent of the Central Province students resist regional
   distribution of resources is to read (implicitly) that 37 per cent
   of the Central Province students favour such distribution.

7. The absence of 100 per cent totals is an important clue in table reading.
   It alerts the reader that one entry is independent of another. When
   a table, or part of a table, totals 100 per cent (as in the occupational
   background table just presented) the entries which total 100 per cent
   are mutually interdependent. (5, p.114)

## Cross-Tabulation Tables

A standard table in social science research shows the relationship
between two variables. Such tables are often called cross-tabulations, because
they simultaneously tabulate the sample on two separate dimensions. They do
this in such a way that the reader can see the inter-relationship between a
respondent's score on one variable (or dimension) and his score on a second
variable.

Let us use a hypothetical study to illustrate several aspects of table
construction and table interpretation. In a small study of rural development
110 farmers are questioned on a variety of issues, including their willingness
to adopt such new farming methods as mechanisation, fertilisers, hybrid seed and
so forth. Out of these questions is constructed a scale of farmer innovative-
ness; and the farmers in the sample are classified into two groups: innovative
farmers and non-innovative farmers. Forty-three of the farmers fall into the
innovative category (39%) and the remaining 67 farmers (61%) fall into the non-
innovative category.

The researcher intends to study possible explanations for farmer innovativeness. In this study, then, the trait of innovativeness or non-innovativeness becomes the dependent variable. Whether a farmer is innovative will be said to <u>depend</u> <u>upon</u> some other condition or trait of the farmer.

The thing upon which innovativeness depends can be called the independent variable. For instance, if innovativeness depends on how close the farmer is to a market road, then 'closeness to a market road' would be the independent variable. If innovativeness depends on whether the farmer grows cash crops, that cash-cropping would be the independent variable.

Normally in table construction the independent variable is placed at the top of the table and the dependent variable down the side of the table. The heading of the table should indicate which variables are being cross-tabulated; we will take cash-cropping as our independent variable.

<u>Relationship Between Cash-Cropping</u>
<u>and Innovativeness of Farmer</u>

Cash-Cropping

| Innovativeness | No | Yes | |
|---|---|---|---|
| No | | | (67) |
| Yes | | | (43) |
| | 100% | 100% | |
| N = | (54) | (56) | |

In cross-tabulating two dichotomous variables a table produces four **cells.** Each cell represents a possible combination of scores on the two variables. Stated differently, each cell represents a possible type of farmer:

| Type of farmer | Whether Farmer Grows Cash-Crops | Whether Farmer Is Innovative |
|---|---|---|
| 1 | No | No |
| 2 | No | Yes |
| 3 | Yes | No |
| 4 | Yes | Yes |

Table analysis is simply the determination of the proportion of units in a study which combine a given pair of values for the variables being cross-tabulated. In the present example, table analysis is the determination of the

proportion of each of the four types of farmers. The stronger the relation-
ship between cash-cropping and innovativeness, the greater the number of
farmers who will be either type 1 (they do not cash-crop, and thus they are
non-innovative) or type 4 (they cash-crop, and thus are innovative).

Most cross-tabulated tables are constructed so that a positive
relationship is demonstrated if most of the cases fall into the diagonal cells,
as shown in the following table:



If the data fall into the diagonal cells as marked, then farmers either have
both of the traits or they lack both of the traits; that is, there is a
positive relationship between the two variables.

If, instead, the data fall into the opposite diagonal, as shown below,
then there is a negative relationship between the two variables. Engaging in
cash-cropping is associated with non-innovativeness.



A third pattern can appear in the table, one suggesting the absence
of any relationship between the two variables: whether a farmer cash-crops bears

no systematic relationship, negative or positive, to whether he is innovative.

In cross-tabulations the table should be percentaged so that the reader can see the effect **on** the dependent variable. Because the **independent** variable appears at the top of our illustrative table, the per cents total at the bottom of the table. Here is the table with data inserted:

|  |  | Cash-Cropping | |  |
|---|---|---|---|---|
| Innovative | No | | Yes | |
| No | 69% | | 54% | (67) |
| Yes | 31% | | 46% | (43) |
|  | 100% | | 100% | |
| N = | (54) | | (56) | |

The table is best read by comparing across the columns; that is, comparing what happens to the dependent variable as the values of the independent variable change. Are farmers who cash-crop more likely to be innovative than farmers who do not cash-crop? If so, there is a positive relationship.

Indeed, the definition of a positive relationship in data analysis is a data pattern in which the values of the dependent variable increase as the values of the **independent variable increase.** In a simple 2 by 2 cross-tabulation such as we are looking at now, **both** the independent and the dependent variable take only two values. It is thus a simple matter to inspect the relationship between the two variables.

The table shows there to be a positive relationship, albeit not a particularly strong one. A higher proportion of farmers (46%) who engage in cash-cropping are innovative than farmers who do not engage in cash-cropping (31%).

The Gamma Measure of Association

Often the researcher wishes to express the relationship of cross-tabulation in the form of a summary statistic. Statistics which measure cross-tabulation relations are often called table statistics or measures of

association. There are very many such statistics, and the researcher normally chooses among them depending on his data and his research questions.

Only one table statistic will be introduced in this text: the gamma measure of association. This statistic illustrates the principle behind table statistics, it is comparatively easy to compute and to interpret and it is more widely used than many other table statistics.

The gamma measure shows the direction and strength of association between two cross-tabulated variables. It varies between -1.00 (indicating a perfect negative relationship) and +1.00 (indicating a perfect positive relationship). A gamma of .00 or near .00 indicates the lack of any relationship between the variables.

The conceptual meaning of the gamma measure is relatively clear. It measures the extent to which the cases are concentrated in one diagonal minus the extent to which they are concentrated in the opposite diagonal. If the cases are equally distributed in each diagonal, then the gamma measure would show a .00 relationship.

Turning to the computation of the gamma measure will help make these ideas clearer. The computational formula is:

$$\text{gamma} = \frac{(A \times D) - (B \times C)}{(A \times D) + (B \times C)}$$

It is convenient to identify the cells in a cross-tabulated table as follows:

| A | B |
|---|---|
| C | D |

It can be seen that the formula for the gamma measure involves the relationship between the A - D diagonal (a concentration of cases here indicates positive relationship) and the B - C diagonal (a concentration of cases here indicates negative relationship).

To compute the gamma measure on the table showing the relationship between cash-cropping and innovativeness we need to insert the actual numbers rather than the proportions, as follows:

Cash-Cropping

| Innovative | No | Yes |
|---|---|---|
| No | 37 | 30 |
| Yes | 17 | 26 |

Applying the formula for the gamma measure to these data:

$$\frac{(37 \times 26) - (30 \times 17)}{(37 \times 26) + (30 \times 17)} = \frac{962 - 510}{962 + 510} = \frac{452}{1472} = .31$$

The gamma measure between cash-cropping and innovativeness for this sample is .31, a positive though moderate relationship.

A few things to note about the gamma measure:

1. If the B - C diagonal were a larger number than the A - D diagonal the numerator would be a negative number. The gamma measure would then be negative, indicating a negative relationship between the two variables.

2. A weakness of the gamma statistic is that it gives a -1.00 or +1.00 if there is an empty cell, as shown below. This weakness should be kept in mind when interpreting the gamma measure.

Cash-Cropping

| Innovative | No | Yes |
|---|---|---|
| No | 54 | 30 |
| Yes | 0 | 26 |

$$\frac{1404 - 0}{1404 + 0} = 1.00$$

Although the illustration we have used is a simple 2 by 2 table, the gamma measure can be applied to much more complex tables. The principle remains the same, but the arithmetic becomes more complicated.

Next is a 3 by 2 table, in which the independent variable has three categories and the dependent variable has two categories. In this case (AxD) is derived by multiplying any given cell frequency by the summed frequencies in all cells which are to the right <u>and</u> lower than the initial cell. We determine (BxC) by reversing the procedure.

To illustrate:

|        | Low | Med. | High |
|--------|-----|------|------|
| Low    | 6   | 11   | 8    |
| High   | 2   | 13   | 17   |

(AxD) = 6 x 30 (30 being the sum of all cells which are to the right
and lower than the initial cell) = 180

11 x 17 (17 being the only frequency which is to the right and
lower than the initial cell) = 187

So, (AxD) for this table is 367. Note that we included as "A" all
cells which had any other cells which were to the right and lower.

(BxC) = 8 x 15 (15 being the sum of all cells which are to the left
and lower than the initial cell) = 120

11 x 2 (2 being the only frequency which is to the left and
lower than the initial cell) = 22

So, (BxC) for this table is 142.

Think back to the 2 by 2 table,**where** we said that the gamma measure showed how
many cases fell along the diagonal which confirmed the hypothesis, as opposed
to those cases which fell along the diagonal which disconfirmed it. You can
see that we have **done** the same thing to this table: the total figure for (AxD)
representing cases which are on the diagonal confirming the hypothesis and the
total figure for (BxC) on the opposite, or disconfirming, diagonal. The actual
measure of association for this table would be:

$$\frac{367 - 142}{367 + 142} = \frac{225}{509} = .44$$

To further illustrate the computation of the gamma measure, we apply
the gamma to a 3 by 3 (nine-cell) table. Again the principle is the same: we
determine the number of cases on the A - D diagonal and the number on the B - C
diagonal. In the example chosen, the relationship is negative, because fewer
cases fall into the A - D diagonal than into the B - C diagonal. The table shows
the relationship between year in university and social class background. A
sample of 93 students are trichotomised on each of the variables and then cross-
tabulated:

Year in University

|  | | 1st | 2nd | 3rd |
|---|---|---|---|---|
| Social Class Background | Low | 6 | 10 | 10 |
| | Med. | 10 | 13 | 13 |
| | High | 14 | 11 | 6 |

$$(AxD) = \begin{array}{l} 6 \times 43 = 258 \\ 10 \times 19 = 190 \\ 10 \times 17 = 170 \\ 13 \times 6 = \underline{78} \\ \phantom{13 \times 6 =} 696 \end{array} \qquad (BxC) = \begin{array}{l} 10 \times 48 = 480 \\ 10 \times 24 = 240 \\ 13 \times 25 = 325 \\ 13 \times 14 = \underline{182} \\ \phantom{13 \times 14 =} 1227 \end{array}$$

$$\text{Gamma} = \frac{696 - 1227}{696 + 1227} = \frac{-531}{1923} = -.28$$

Correlation Matrix

For the reader of social science reports who is untrained in social statistics, one of the most daunting types of tables is the correlation matrix. Actually such tables are easy to read, once it is clear what it is the author is trying to accomplish. A correlation matrix simply presents in summary fashion the relationships between a large number of pairs of variables. This can be shown through example.

Dr. Philip Mbithi in his studies of rural development in Kenya has investigated the degree to which systematic inter-relationships characterise a wide variety of farmer behaviours, attitudes and practices. In one research report he writes about a

> high degree of intercorrelation between such diverse scales as farm mechanization, crop husbandry practices, household level of living (measured by items such as radios and books), social participation (measured by membership in various formal and informal organizations such as churches, co-operatives and self-help groups), and formal contacts (with Government agents, in particular).

Finding, as he does, a pattern of intercorrelation, Mbithi concludes: "Development activity, like all basic human enterprises, is patterned and systematic and is not characterized by erratic shifts, discontinuities and inconsistencies." (16, p. 77)

Data presented to support this assertion include the following table.

Intercorrelations of Differentiation Scales and Farm Operation Measures, East Kenya (16, p. 78)

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. Scale of mechanization | — | .27 | .71 | .42 | .37 | .69 | .86 | .76 | .64 | .58 | .25 | .53 | .27 | .11 | .48 | .65 |
| 2. Scale of crop husbandry | | — | .27 | .44 | .48 | .62 | .31 | .94 | .56 | .31 | .36 | .30 | .54 | .23 | .59 | .81 |
| 3. Scale of level of living | | | — | .34 | .32 | .42 | .51 | .64 | .34 | .26 | .27 | .40 | .01 | .11 | .06 | .30 |
| 4. Scale of formal contacts | | | | — | .41 | .60 | .69 | .84 | .42 | .56 | .29 | .40 | .21 | .00 | .12 | .22 |
| 5. Scale of social participation | | | | | — | .39 | .49 | .72 | .12 | .64 | .19 | .01 | .12 | .42 | .34 | .22 |
| 6. Scale of farming standard | | | | | | — | .63 | .88 | .32 | .46 | .27 | .69 | .66 | .32 | .42 | .48 |
| 7. Scale of animal husbandry | | | | | | | — | .80 | .26 | .38 | .31 | .43 | .06 | .29 | .29 | .21 |
| 8. Farm adoption score | | | | | | | | — | .43 | .26 | .13 | .52 | .33 | .38 | .47 | .49 |
| 9. Non-farm occupation | | | | | | | | | — | .42 | .39 | .08 | .12 | .15 | .24 | .40 |
| 10. Geographic mobility | | | | | | | | | | — | .22 | .01 | .01 | .21 | .03 | .28 |
| 11. Education | | | | | | | | | | | — | .10 | .26 | .26 | .10 | .26 |
| 12. Farm income | | | | | | | | | | | | — | .38 | .26 | .16 | .47 |
| 13. Score of farm constraints | | | | | | | | | | | | | — | .14 | .46 | .32 |
| 14. Agricultural training | | | | | | | | | | | | | | — | .09 | .28 |
| 15. Farm planning and rotation | | | | | | | | | | | | | | | — | .41 |
| 16. Cash-crop ratio | | | | | | | | | | | | | | | | — |

Total cases = 126 farmers.
General level of significance P = 0.05, gamma = .32.

The measures of association or correlation reported in this table are gammas, exactly of the sort we have just reviewed, and the table is easily read and interpreted. Each one of the sixteen variables listed down the left-hand side of the table is correlated with all of the remaining fifteen variables. We can see that the table can answer very many questions for us:

1. Do farmers who practice mechanised farming have a higher standard of living than farmers who do not practice mechanised farming?

   Yes, there is a strong positive association (.71) between the measure of mechanisation and a measure of living standards.

   To see this the reader need only find the intersection between variable 1 (measure of mechanisation) and variable 3 (measure of level of living).

2. Do farmers who have received agricultural training have more formal contacts than farmers who have not received such training?

   No, not at all, there is no relationship between these two variables. (see the intersection between variable 14 and variable 4. )

3. Which is more strongly related to a farmer's income, his formal education or his use of cash-crops?

   Cash-cropping appears to increase farmer income more effectively than does his formal education; the association between cash-cropping and income (.47) is nearly five times as strong as the association between education and income (.10).

In reading a matrix of correlation measures such as that presented by Mbithi, one need only understand that each entry in the table is summarising a standard cross-tabulation. When one sees a particular gamma measure then, it is useful to also see in the mind's eye the type of table from which the gamma measure might have been constructed. A gamma as low as .10 (as between education and income) tells the reader that the proportion of farmers in this sample with high income does not significantly increase from one education category to another.

Here, for instance, is a hypothetical table showing a relationship between education and income which would produce a gamma of .10, based on 126 cases.

### Relationship Between Farmer's Education and His Farm Income

| Farm Income | Level of Education | | |
|---|---|---|---|
|  | Low | Moderate | High |
| Low | 54% | 50% | 46% |
| High | 46% | 50% | 54% |
|  | 100% | 100% | 100% |
| n = | (50) | (50) | (26) |

Developing a knack for visualising the strength of relationships, expressed in proportions and summarised by any particular level of gamma, can greatly facilitate correct interpretation of a matrix which presents a large number of relationships in summary fashion.

CHAPTER 8

THE MEAN AND THE STANDARD DEVIATION

The result of a typical sample survey is a large number of variables which have been measured for each unit in the survey.  In a rural development survey, for instance, each household will have been measured along dozens of such variables as:

Size:     the number of members in each household;

Crops:    the amount of land in production for each crop farmed;

Health:   the frequency of particular illnesses among household members;

Community  the number of self-help projects in which household members
Involvement:  have actively cooperated; and

Political  the number of media sources regularly used by household
Information:  members to learn about the affairs of their society.

Even for such a partial list as these five variables, no reader can begin to look at the score for each specific variable for each individual household.  Besides, the scores for individual households do not mean much.  They begin to take on meaning only when compared with the scores of other households.

The initial research task is to reduce the mass of individual scores to some manageable summary scores.  This is one purpose of statistics.  Statistics summarise a large number of scores so that the researcher can quickly see what the general characteristics of the sample are.  In this chapter we review two of the most basic descriptive statistics in   the social sciences: the mean and the standard deviation.

The mean is often called a measure of central tendency, for it tells the researcher about the main or central characteristics of a distribution of scores.  And the standard deviation is called a measure of dispersion, for it tells the researcher how the scores are spread out or dispersed.  The notions of central tendency and dispersion will occupy our attention in this chapter, as we review in some detail the mean and standard deviation.

This detailed review has two purposes: first, the statistics described are useful tools in their own right, and thus are worth a detailed understanding; second, both the mean and the standard deviation play important roles in correlation

statistics described in the next chapter. The present discussion, then, is background for consideration of more advanced statistics.

We begin with some basic statistical notation:

N = Number, or the number of units on which a statistic is being calculated. Usually the 'N' is the total number of cases in a sample.

x = An individual score or measure, such as the size of a particular household or the participation level of an individual citizen.

$\bar{X}$ = The mean, or arithmetic average.

s.d.= An abbreviation for the standard deviation. (Often in statistical textbooks or research reports the standard deviation is symbolised by the Greek letter sigma,$\sigma$, but it is just as easy to use the roman letters, and that will be the practice in this text.)

$\sqrt{\phantom{xxx}}$ = The Square root.

$\Sigma$ = This is the Greek capital letter sigma, which is used to indicate a summing operation. When $\Sigma$ appears in statistical notation, the reader is being instructed to sum whatever follows.

For instance, the notation $\Sigma x$ is an instruction to sum all the individual scores.

## The Mean or Arithmetic Average

Although the mean and its computation are familiar, some of the properties of the mean are less familiar. A review is useful.

The mean is obtained by summing the individual values and dividing by the total number, thus $\bar{X} = \dfrac{\Sigma x}{N}$ , where x stands for individuals scores.

It is the typical score or value in a series of scores or values. Here are five scores obtained on an examination: 72, 81, 86, 69, 57. The mean or typical score is 73. One way of understanding a typical score is to think of it as that score which every student would have received if the total of all five exams had been distributed evely. Under conditions of equality of distribution, everyone receives the mean score or value.

1) When the mean is multiplied by the total N, the result is the sum of the original distribution of scores: 73 x 5 = 365, which is the sum of the five figures, 72, 81, 86, 69, 57. This property

of the mean is important in using a small sample to estimate
the total of a much larger group.  To estimate the total yield
of a large coffee farm, we take a sample of just a few coffee
plants, and calculate the average (mean) yield of those plants.
We multiply this figure by the number of coffee plants on the
farm and have an estimate of the total yield.

2) Another important property of the mean is that the sum of the
deviations from the mean is zero.  Stated differently, the
total amount by which some scores are above the mean (positive
values) is exactly equal to the total amount by which the
remaining values are below the mean (negative values).  Thus,

$$
\begin{array}{cccc}
& X & \bar{X} & X-\bar{X} \\
\text{Scores above} & 57 & 73 & -16 \\
\text{the mean} & 59 & 73 & -4 \quad = -21 \\
& 72 & 73 & -1 \\
\text{Scores below} & 81 & 73 & +8 \\
\text{the mean} & 86 & 73 & +13 \quad = \dfrac{+21}{0} \\
& \overline{365} & &
\end{array}
$$

One can imagine each score in a distribution of **scores as a weight**.
The mean, then, is the point of equilibrium, the point where the weight of
scores on one side is balanced by the weight of scores on the other side.  In
the example just cited, the sum of the positive deviations, the weights on one
side of the mean, is equal to the sum of the negative deviations, the weights
on the other side of the mean.  It is in this sense that the mean can be thought
of as the point of equilibrium.

3) A further important property is that the sum of the squared
deviations of each score from the mean is a minimum.  That is,
it is less than the sum of the squared deviations from any other
number.  In the distribution of exam scores listed above the
sum of the squared deviations is 506.  This is a number smaller
than the number one would get by using any constant number other
than the mean from which to calculate the deviations.  This
property of the mean is basic to correlation and regression
analysis, and in that context is called the least squares principle.

The least squares principle is easily proved.  In the above
distribution the five deviations from the mean were:

| deviations | deviation squared |
|---|---|
| - 16 | 256 |
| - 4 | 16 |
| - 1 | 1 |
| 8 | 64 |
| 13 | 169 |

sum of squared deviations  =  506

Now, instead of using the mean (73) we use another number and cal-
culate the sum of the squared deviations.  Even if the number chosen is
practically the same as the mean, the sum will be larger.  Here are the figures
using 72 instead of 73

| score | | | | deviation | deviation squared |
|---|---|---|---|---|---|
| 57 | - | 72 | = | -15 | 225 |
| 69 | - | 72 | = | - 3 | 9 |
| 72 | - | 72 | = | 0 | 0 |
| 81 | - | 72 | = | + 9 | 81 |
| 86 | - | 72 | = | +14 | 196 |

sum of squared deviation      =      511

The mean is an extremely useful statistic.  It quickly summarises
a large amount of information.  The mean rainfall of a region tells a crop
scientist the type of crops likely to grow in that region.  The mean exam
score of a secondary school tells an educator how effectively that school
is preparing its students.  The mean weight and height of a particular age
group tells the medical scientist whether essential nutrition standards are
being met.  The mean number of cattle owned by a group of nomadic families
gives the agricultural economist an insight into rural poverty.

Means can easily be compared, which is another strength of this
statistic.  Comparing mean exam scores across a large number of schools indicates
how schools are doing in comparision with each other.  Comparing the mean rain-
fall at different elevations gives an idea of what crops might be effectively
introduced into what regions.  In a later chapter we also review a statistical
procedure useful for interpreting the difference between means.

## Measures of Dispersion

The mean has one important weakness not yet noted. Very different distributions of scores can give the same mean. The mean for a set of five exam scores - 57, 69, 72, 81, 86 - was 73. Below are listed two additional distributions of five exam scores, both of which also provide a mean of 73.

| | |
|---|---|
| 43 | 71 |
| 63 | 72 |
| 78 | 73 |
| 83 | 74 |
| 98 | 75 |
| 365 | 365 |
| $\bar{X} = 73$ | $\bar{X} = 73$ |

Although the first group of students has the same average performance as the second group, the two groups are nevertheless very different. Exam performance is highly variable in the first group and very homogeneous within the second. Perhaps the teacher of the first group spends all of his energies preparing the two best students for the exam and largely **ignores** the poorer students, whereas in the second school an effort is made to bring everyone up to the average performance level.

Whatever the explanation, it is obvious that the mean can be a misleading statistic. It tells nothing about the distribution of scores, or what technically is called the dispersion around the mean. This is the chief limitation of the mean as a summary statistic. Consider another example. Take two hypothetical countries, each with an average annual rainfall of 60 inches. Knowing this, one might be tempted to conclude that the basic agriculture of these countries will be similar. This is a conclusion to be resisted until something is known about the regional variation within each country. Perhaps the regional distribution in the two countries is something such as follows:

| Country A | $\bar{X}$ rainfall | Country B | $\bar{X}$ rainfall |
|---|---|---|---|
| North | 10 inches | North | 45 inches |
| Central | 70 inches | Central | 65 inches |
| South | 100 inches | South | 70 inches |
| Overall $\bar{X}$ = 60 inches | | Overall $\bar{X}$ = 60 inches | |

Clearly agricultural practices will differ enormously from one country to the other, despite the fact that both countries have the same average annual rainfall.

What is needed is a measure of distribution or dispersion to go along with the mean. That is, any set of scores can be characterised in two major ways: its average or typical score (the mean) and its dispersion around the mean.

There is more than one statistic which provides information about dispersion. The simplest is the range, which is the difference between the highest and lowest score. Our initial distribution of exam scores had a range of 29 (the difference between 86, the highest exam score and 57, the lowest). The other two distributions had ranges of 55 (98 - 43) and 4 (75 - 71). We could conclude, correctly, that the set of scores with a range of 55 was the most dispersed set and the set with a range of 4 was the least dispersed.

However, the range is not always a very useful statistic. It is based on only two scores, the most extreme two. A more useful statistic is one which takes account of every score in the distribution. The most versatile and commonly used summary statistic measuring dispersion is the standard deviation.

## The Standard Deviation

A standard deviation is a statistic which measures dispersion (or deviation). In any set of scores, each individual score deviates from the mean by some amount (if a score is identical to the mean, we say that it has a deviation of zero). The standard deviation calculates each individual deviation, and then summarises these in such a way as to tell us whether the scores are bunched around the mean or widely scattered around the mean. (The term "standard" has a special meaning, considered below.)

Deviation, then, refers to how much an individual score is greater than or less than the mean of the set of scores.

The steps in calculating a standard deviation are:

a. calculate the mean

b. obtain the difference (deviation) of each individual score from the mean

c. square the deviations

d. sum the deviations and divide by n (the total number of individual scores).

The figure obtained as the result of these operations is the variance.  The square root of this figure is the standard deviation.  The formula for standard deviation is:

$$\text{standard deviation} = \sqrt{\frac{\Sigma(\bar{X} - x)^2}{N}}$$

Here is the computation for the distribution of exam scores noted above:

| x | $(\bar{X} - x)$ | $(\bar{X} - x)^2$ | |
|---|---|---|---|
| 72 | 1 | 1 | |
| 81 | -8 | 64 | s.d. $= \sqrt{506/5} = \sqrt{101.2} = 10.06$ |
| 86 | -13 | 169 | |
| 69 | 4 | 16 | |
| 57 | 16 | 256 | |
| | 0 | 506 | |

$\bar{X} = 73$

Take note that the greater the spread of scores around the mean, the larger the standard deviation.  Had all five scores been 73 (that is, been identical and therefore equal to the mean) then the deviations would have been zero and so also would have been the standard deviation.  Note also that extreme values count heavily in computing the sum of the deviations, which can sometimes give a mis-leading impression.  In the above example, the number 101.2 is the variance, a measure of importance when we discuss correlation.

One extremely important idea is sometimes not fully appreciated.  Ask yourself, before reading on, what a standard deviation is a measure of.  For instance, in the example above, what does the 10.06 stand for?

The 10.06 is of course related to exam scores.  An exam score which is one standard deviation above the mean would be an exam score of 83.06, just as an exam score one standard deviation less than the mean would be a score of 62.94, though we would round these scores to 83 and 63.

A standard deviation is always a measure expressed in the same terms as the original scores.  If the original scores are inches of rainfall, then the standard deviation is a measure of inches of rainfall.  If the original scores

are crop yield, then so is the standard deviation.  If the original scores
are levels of political participation, then so is the standard deviation.
Another way of thinking about this is to realise that whatever the mean is
a mean of the standard deviation is a standard deviation of.

## The Normal Curve

We have asserted that the standard deviation is a very useful
statistic, but have not yet explained just how it is useful.  What does it
tell us to say that the standard deviation of a set of exam scores is 10.06?

To answer this question we introduce the idea of the normal curve
(sometimes called a Gaussian curve, or a bell-shaped curve).  A normal curve
occurs when in a distribution of scores there are a few very low scores, a
few very high scores, a few more moderately low scores, a few more moderately
high scores, and most scores clustered fairly close to the mean.  Here is a
hypothetical distribution of exam scores reproduced as a normal curve.



No. of Persons
with each score

```
lowest          mean          highest
scores          score         scores
```
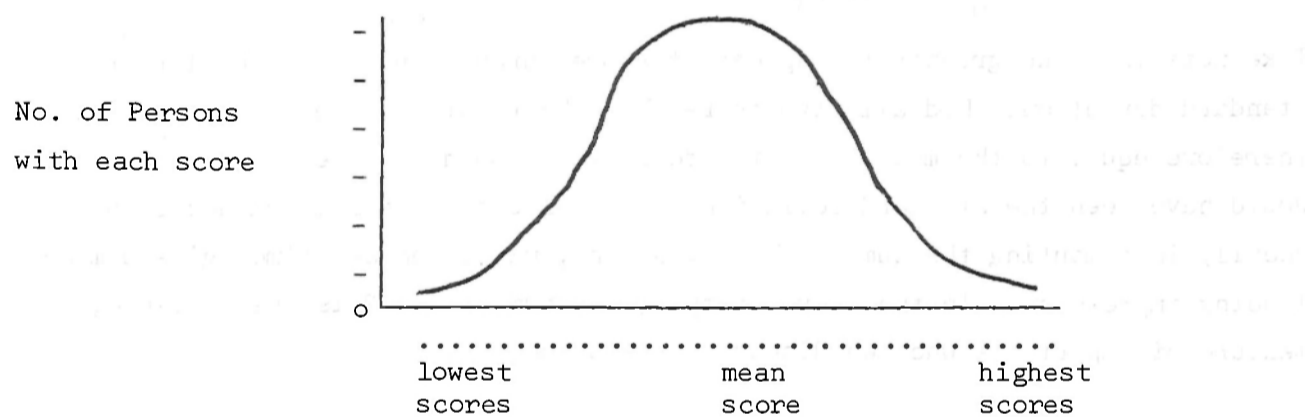
Very few persons get the lowest score, and very few get the highest score.
The largest number of persons get a score fairly close to the mean.

This profile or distribution of scores is a normal distribution.
Of course scores need not be normal, and when they are not we speak of skewed
distributions.  Such is the case for income, for instance; income distributions
do not cluster around the mean, but more often show a very, very few people
with the higher incomes and a large proportion of the population with low
incomes.

What is remarkable, however, is how many things in the world are
more or less normally distributed.  In a forest, for example, there are a few
very large trees and few very small trees, but most trees are about average

in height.  In a library there will be a few very fat books and a few very
thin ones, but most books will be about average in size (prove this to your-
self next time you visit the library).  Intelligence is, the psychologists
tell us, fairly normally distributed; there are not very many geniuses or very
many idiots (though a few of both) but a good many about average persons.
A football team will have one or two really superior players, a few who only
barely manage to stay on the team, and many more players who are about average.
Over a fifty-year time span, there may be a few years of excessive rainfall
and few years of very meagre rainfall; most years will not be extremes but
will be, again, about average.

The idea of normal distributions is well worth a few minutes reflection.
Think up other examples of things which are probably normally distributed.  You
will be impressed by how many things you come up with.  Then try to think of
things which most likely are not normally distributed (income is one example
we gave earlier; another might be the nuclear capabilities of the nations of
the world).  You might want to ponder **why**  so many things are normally distributed;
and also why there are exceptions.  You will notice that things which are given
"by nature" are more likely to be normally distributed than are things which
result from the control of political or military power.

The connection between the normal curve and the standard deviation is
what interests us here.  The essential fact is simple.  In a normal distribution
slightly more than two-thirds of the scores (68.26% to be exact) will be within
plus or minus one standard deviation of the mean.  No other single fact is as
important as this one is when it comes to understanding the statistics reviewed
in this text.

A sample of 1,000 farmers has been asked how far their farm is from
the nearest marketing cooperative.  The **answers** range from a low of 3 kilometers
to a high of 47 kilometers for the few very isolated farmers.  The study
reveals that:

> The mean distance away from a cooperative is 25 kilometers.
>
> The standard deviation is 7 kilometers.
>
> The distribution is normal; a few farmers are very close to
> a marketing cooperative; a few are very far away; most are
> neither very close nor very far.

In a normal distribution, approximately two-thirds of the cases will be within plus or minus one standard deviation of the mean. More precisely, one-third of the cases will be within one standard deviation above the mean and one-third will be within one standard deviation below the mean.

It follows that two-third of the farms are between 18 and 32 kilometers of a marketing cooperative. How do we know this?

18 equals the mean minus one standard deviation
(25 - 7 = 18)

32 equals the mean plus one standard deviation
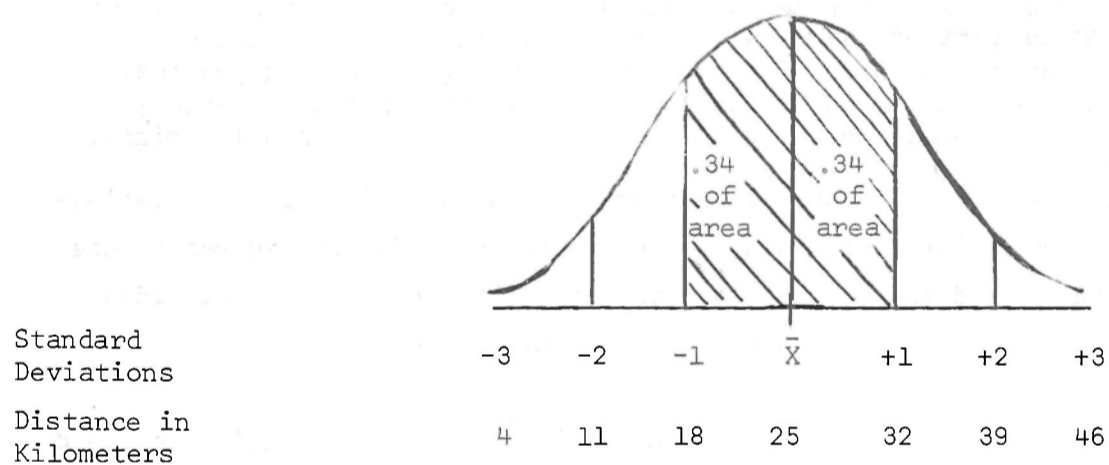(25 + 7 = 32)

Note that all these numbers stand for kilometers, which is the thing being measured. The standard deviation tells us the proportion of farmers who live a given distance away from a marketing cooperative, and herein lies the enormous utility of this simple summary statistic.

How far away from a marketing cooperative does a farmer live if he is two standard deviations away from the mean? You should note that the question has been poorly put, because it does not specify if the farmer is above or below the mean. If he is above the mean, he lives 39 kilometers away; if he is below the mean, he lives 11 kilometers away. Why? Because two standard deviations is 2 x 7, or 14 kilometers. Satisfy yourself that two standard deviations **above** the mean represents 39 kilometers; and that two standard deviations below **the** mean represents 11. Next determine how far away a farmer is who is three standard deviations above the mean, and how far one is who is three standard deviations below the mean. There is a reason for considering two and even three standard deviations away. For just as the normal curve indicates the proportion of farmers within one standard deviation, it indicates the proportion within two or three standard deviations.

Approximately 95 percent (95.46%) of all the farmers will be within two standard deviations. Nearly all of the farmers (all but 3 out of 1,000) will be within three standard deviations. These proportions, of course, presume a normal distribution. Here is the full distribution:

| | Standard<br>Deviation | Proportion<br>of Farmers | Distance From<br>Marketing Cooperative |
|---|---|---|---|
| Below the<br>mean | -3<br>-2<br>-1<br> | .1<br>2.4<br>13.5<br>34.0 | less than 4 kilometers<br>between 4 and 11 kilometers<br>between 11 and 18 kilometers<br>between 18 and 25 kilometers |
| X̄ | | | |
| Above the<br>mean | +1<br>+2<br>+3<br> | 34.0<br>13.5<br>2.4<br>.1 | between 25 and 32 kilometers<br>between 32 and 39 kilometers<br>between 39 and 46 kilometers<br>more than 46 kilometers |

When we say that such and such a proportion of the cases are within so many standard deviations, we are using language to describe what is often called the area under a normal curve. The proportion of farmers living various distances from the marketing cooperative is equivalent to the area of the normal curve which falls within a given standard deviation from the mean.



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Standard<br>Deviations | -3 | -2 | -1 | X̄ | +1 | +2 | +3 |
| Distance in<br>Kilometers | 4 | 11 | 18 | 25 | 32 | 39 | 46 |

The lined area under the normal curve represents approximately two-thirds of the total area covered by the curve. It represents as well the scores of approximately two-thirds of the farmers.

It is possible to measure distances from the mean in other than exact multiples of the standard deviation. For example, 99 per cent of the cases in a normal distribution fall within plus or minus 2.58 standard deviations of the mean. In the present example;

$2.58 \times 7 = 18.06$; thus 99 per cent of the farms are within about 18 kilometers of the mean for the distribution, or are between 43 and 7 kilometers away from a marketing cooperative. Only one per cent of the farmers live nearer than 7 kilometers or further than 43 kilometers.

## The Normal Curve and Sampling Theory

At the conclusion of the chapter on sampling theory we introduced the notion of confidence interval or confidence level. A researcher is 95 per cent confident or 99 per cent confident that his sample accurately estimates the characteristic of the population from which it is taken.

Confidence intervals are closely related to the properties of the normal distribution. We can repeat part of the answer given by the researcher when asked what it means to say that he is "95 per cent certain that his sample estimate reflects the population characteristic ".

> This 95 per cent confidence level means the following. If I had taken 100 samples instead of only one sample, then in 95 of these samples we would have found between 28 and 32 per cent of the farmers to be fertiliser adopters. However, in five of these samples we would have found the number of adopters to be either less than 28 per cent or more than 32 per cent.
>
> Of course, I did not take 100 samples. I took only one. And I cannot know for certain whether the one I took is part of the 95 or part of the remaining five. If the sample actually drawn is part of the five, then I'm wrong in telling you that the population adoption rate is within the specified range. If the sample actually drawn is part of the 95, then I'm right.

The researcher is here reflecting his understanding of the relationship between normal distributions and sampling theory. The procedures he uses are fairly complex and are not reviewed in this text. But the central idea can be grasped with an understanding of the normal curve.

The connection between sampling statistics and the material reviewed in this chapter derives from notions of probability. Probability is in turn directly related to the idea of area under the normal curve.

If we choose at random an individual farmer from our sample and ask what is the probability that he lives more than 25 kilometers away from a marketing cooperative, we would quickly answer that the probability is .5. After all, half, or .5, of the farms are further away than 25 kilometers, and thus the chances that any give farmer lives this far away is also .5.

To use different language, the probability is .5 because .5 of the area under the normal curve lies above the score of 25 kilometers. What is the probability that the randomly chosen farmer lives within 11 kilometers of

a cooperative?  The probability is .025, because .025 of the normal curve
is represented by the area 11 kilometers or less from a cooperative.

What is the probability that a farm is either less than 11 kilometers
away or more than 39 kilometers away?  Now we have to add two areas of the
normal curve, the area less than 11 kilometers and more than 39 kilometers, or
.025  +  .025 = .05.  That is, the chance that a farm is closer than 11 kilo-
meters or further than 39 kilometers is 5 per cent.

The last example can be reversed: what is the probability that the
farm is not closer than 11 kilometers or further than 39 kilometers?  Clearly
the answer is .95, because 95 per cent of the farms are within this range
(just as 95 per cent of the area of the normal curve is between the score of
11 and the score of 39).

With these examples in mind we return to the idea of sampling and
confidence intervals.  When the researcher declares himself to be 95 per cent
confident that the sample mean reflects the true population mean (within the
range established by the sample error) he has used sampling statistics to
calculate this probability.  The procedures are analogous to calculating the
probability that a randomly selected farm is a given distance away from the
mean of all the farms.

## An Illustrative Application of the Standard Deviation

The weakness of the mean as a summary statistic is the strength of
the standard deviation, for whereas the former tells us nothing about the
dispersion of scores the latter tells us a great deal.  Here is a hypothetical
instance in which knowledge of the mean without information about the dispersion
of scores led to poor public policy.

A planning officer in the Ministry of Agriculture recently attended
an international conference and learned of a hybrid maize seed which has
excellent yield in nearly any soil, as long as the rainfall is at least 60
inches during the growing season.  If the rainfall drops much below 60 inches,
the yield is poor from this seed; indeed, it is much less than the yield produced
by the seed now in common use.

The planning officer intends to introduce the hybrid seed in a demonstration area of the country, but cannot convince farmers to plant it unless he guarantees them an income from maize equal to that which they are presently receiving.  This guarantee will cost the Ministry a considerable sum if yield is much lower than average, and the planning officer has to make many promises to the permanent secretary who authorises the programme.

In preparing for the demonstration the agricultural officer checks the rainfall tables and, wanting to be cautious, chooses an area in which the average rainfall over the past ten years has been 63 inches, three inches more than is necessary to make the hybrid seed a success.

The programme is set in motion.  In the first year the hybrid seed is a great success (rainfall that year was about average, 63 inches).  The officer hears talk that he is under consideration for a major promotion. The second year the rainfall is only 58 inches; the maize crop in the demonstration area is a disaster, a costly disaster for the Ministry of Agriculture which must request a special supplement from Parliament.  Talk about promoting the planning officer evaporates.  And the farmers in the area are reluctant to cooperate with other special Ministry programmes.

What went wrong?  A close look at the rainfall table shows that the area chosen had an unusually high mean rainfall, but also an unusually high standard deviation.  For the ten-year period the standard deviation was 5 inches, that is, there is enormous variability in rainfall in the area chosen for the demonstration.  With a mean of 63 inches and a standard deviation of 5 inches every **sixth** year or so the rainfall will be as low or lower than 58 inches. (Fifty-eight inches is one standard deviation below the mean; about one-**sixth** of the normal curve is represented by the area more than one standard deviation below the mean.)

A wiser choice for a demonstration would have been an area with a much lower standard deviation even if it had a lower mean rainfall as well. A painful lesson has been learned.

CHAPTER 9

STATISTICAL CORRELATION

The mean and the standard deviation describe important properties
of a single distribution of scores, such as the distance of a sample of farms
from a marketing cooperative.  Social scientists find many uses for statistics
which describe distributions, but seldom do social scientists stop at this
point.  Social science is in large measure a search for relationships between
two (or more) variables.  Do farmers who live near marketing cooperatives
realise a higher profit for their crops than farmers who live far away from
cooperatives and thus must subtract transport costs from profits?  A study
designed to answer this question would correlate a measure of distance from
cooperatives with a measure of farm profits.

It is the ideas of correlation that we take up in this chapter.
The material to be covered depends on a prior grasp of the mean and the standard
deviation.  Thus the more thoroughly the reader understands the previous
chapter, the easier will be the material in this chapter.  And if the essential
properties of the mean and the standard deviation are not understood, there is
no way to really see what statistical correlation is all about.

We start with a few simple definitions, which need not be memorised
but which provide an overview of the central ideas in this chapter.

Definitions

A simple correlation coefficient measures the direction and the strength
of relationship between two variables.  (Multiple correlational analysis deals
simultaneously with more than two variables, and is not discussed in this text.)
The particular statistic described in this chapter is known as the Pearson
correlation coefficient, named after the statistician who invented it.

This correlation coefficient tells us whether two variables are related
across a sample of units.  It indicates whether the relationship is positive
or negative, and whether it is weak or strong.  Take age and height.  If among
a sample of students the older tend also to be taller we say that age and
height are positively correlated.  If the tallest student were also the oldest,
and the next tallest the next oldest, and so on until we came to the shortest
student and he was the youngest, then the correlation between age and height

would be perfect or +1.00. Seldom are variables perfectly correlated, however. The Pearson correlation coefficient can take a value from -1.00 (perfect negative correlation) through 0 (no relationship between the variables) to +1.00 (perfect positive correlation). Thus, a correlation of +.85 would indicate a positive and strong, though not perfect, correlation.

Regression analysis is used to predict the exact value of one variable from knowledge of another variable. It is a more demanding (and more valuable) technique than is correlation analysis, though regression and correlation are closely linked. Return to the example of height and age among a classroom of students. Correlation analysis only tells us how strongly the two variables are related. Regression analysis would tell us something more precisely about the nature of the relationship. For instance, can we predict the height of a given student from knowledge of his age? If 9-year-old students are always 48 inches tall, then we could with information about age predict height (at least of 9-year-olds). If each age had associated with it a given height, then we would predict height exactly for the entire class of students from knowledge of age.

Matters are a bit more complicated than this example indicates, however. Regression analysis asks how much an increment in one variable produces an increment in another (dependent) variable. If children always grew exactly six inches a year, we would say that one year of age produces six inches of growth. But of course some children grow less and some children grow more, and thus the prediction will have errors in it. If errors are too great, the prediction is useless. Such of course would be the case if age and height were not correlated. One must find a high degree of correlation between variables before any useful predictions can be made.

The Goal of this Chapter

The actual computation in correlation and regression analysis is extensive (and is usually done for the researcher by a computer). We are not in this chapter going to present all of the statistical formulae, and we are not going to learn the actual steps of computation. Our goal is to attempt to help the reader understand what is going on in correlation and regression analysis.

This understanding can be advanced with verbal descriptions and with illustrative graphs. But it is important to realise that we are not explaining all the steps which a researcher would take to compute a correlation statistic or draw a regression line. We do not answer the question,'How does one know where to draw the regression line?' but rather only,'What is meant by a regression line and a correlation coefficient?'

### An Example

Assume that the Ministry of Agriculture has distributed hybrid maize seed to a sample of 250 farmers scattered throughout the country. The annual maize yield of each of these farmers is measured before the hybrid seed is introduced and then measured one year later. This second measure shows that every user somewhat increased his yield, but that some farmers increased their yields at much greater rates than other farmers. The average yield increase was 40 per cent, but some farmers increased yields as much as 70 per cent and others as little as 10 per cent. The research shows a mean increase of 40 per cent and a standard deviation of 12 per cent.

The Distribution of the Dependent Variable: The amount of yield increase following introduction of hybrid seed is the dependent variable. It is the thing to be explained. More exactly, we want to explain the variability in yield increase. Why do some farmers do so well with the hybrid seed, and others so poorly? Our dependent variable is normally distributed. This should immediately tell you that the distribution is as follows:

>      2 per cent of the farms increased productivity less than 16 per cent;
>     14 per cent increased productivity between 16 and 28 per cent;
>     34 per cent increased productivity between 28 and 40 per cent;
>     34 per cent increased productivity between 40 and 52 per cent;
>     14 per cent increased productivity between 52 and 64 per cent; and
>      2 per cent increased productivity more than 64 per cent.

It will help later if we remind ourselves that this distribution can be plotted. In the following bell-shaped curve, each dot represents a farmer. We see that a few farmers increased yield very little, a few increased yield considerably, and the remainder are between the extremes, with the heaviest concentration being about average, or near the mean.

Number of
Farmers
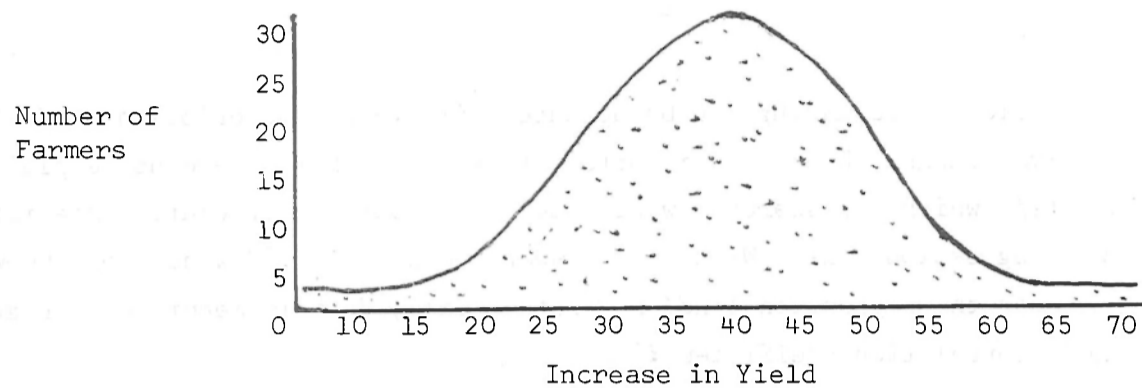


Increase in Yield

The Independent Variable:  The research staff of the Ministry of Agriculture is
puzzled by the effect of the hybrid maize seed.  If they could understand the
variability in yield increase they might be able to take necessary action to
bring every farmer to the level of the highest producers.

Of course very many different things might cause the variability.
Perhaps the extension staff in certain areas did not explain the new seed
correctly, and thus the farmers in that area did not plant according to
instruction.  Or perhaps the hybrid seed works under some soil conditions more
effectively than under other conditions.  If so, continued experimentation
with the hybrid seed might improve its usefulness under all soil conditions or
might lead to a fertiliser which could compensate for whatever is lacking in
the soil.  Yet another factor which might explain the variability is rainfall.
Perhaps the seed works better in high rainfall areas than in low rainfall
areas, and thus what would be needed would be an irrigation programme combined
with the hybrid seed.

The research staff decides to concentrate on the third of these
possible factors.  They plan to correlate the measure of farmer productivity
(how much yield increase is achieved with the hybrid seed) with inches of rain-
fall in each area.

From this point on we refer to increase in productivity as the dependent
variable and inches of rainfall as the independent variable.  Commonly used
statistical notation refers to the dependent variable as the Y variable, and to
the independent variable as the X variable, a usage adopted here.

Scattergram:  Each farm is now measured on two dimensions: how much it increased
productivity (Y variable) and how much rainfall there is in its area (X variable).

A scattergram plots the intersection for each farm of its two scores. The scattegram is a two-dimensional space produced when the Y variable and the X variable are coordinated. Four farms with the following scores are plotted on the scattegram.

Y axis

Percentage of
Productivity
Increase

X axis

Inches of Rainfall
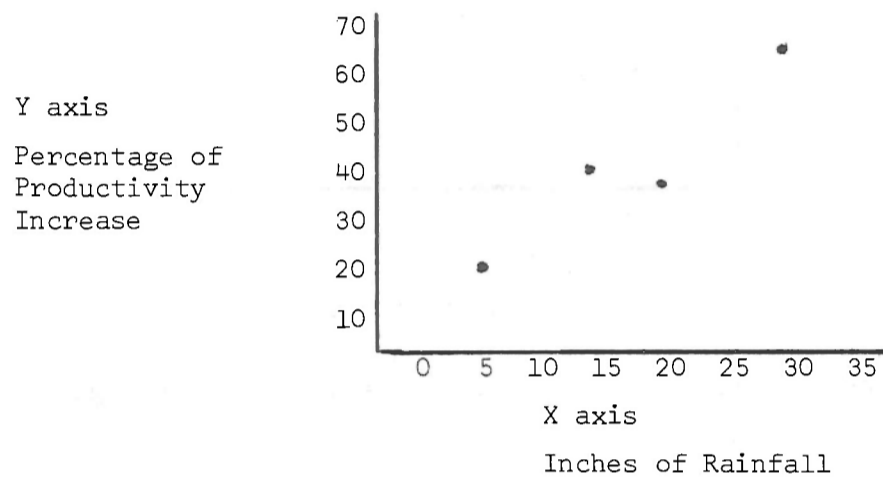
| Y | X |
|------|-----------|
| 20% | 5 inches |
| 65% | 30 inches |
| 40% | 15 inches |
| 35% | 20 inches |

An actual scattergram used with the sample of 250 farms would plot the intersection of every farm's Y score with its X score. Visual inspection of a scattergram can give one an idea whether the two variables are related; and if they are related, whether the relationship is positive or negative. Consider three hypothetical scattergrams. The first one indicates no relationship between the variables; the score on Y (our dependent variable) does not vary with, or depend upon, changes in the score on X (our independent variable).

Y axis

Percentage of
Productivity
Increase

X axis
Inches of Rainfall
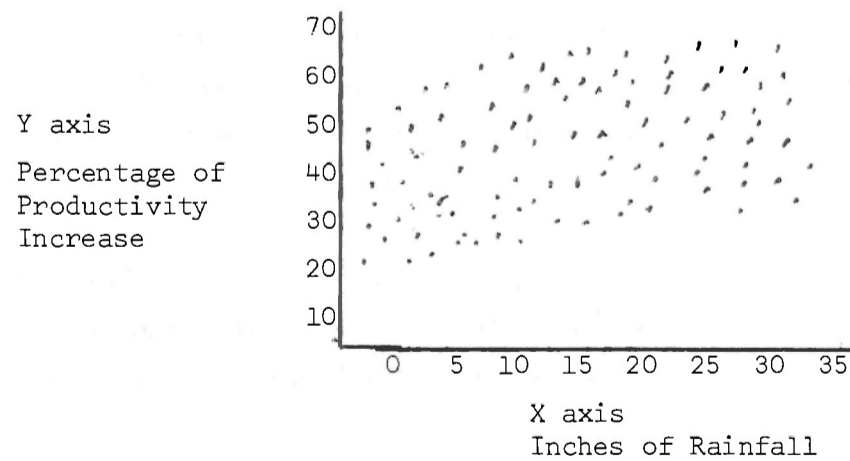
In the next hypothetical scattergram, the relationship is positive.
Yield appears to increase as rainfall increases.

Y axis

Percentage of
Productivity
Increase



X axis

Inches of Rainfall
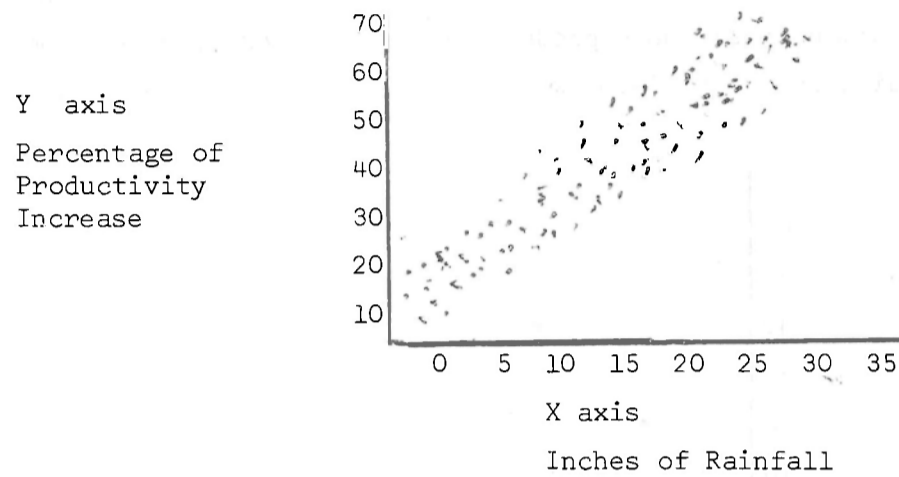
In the last hypothetical scattergram, the relationship is negative.
Yield appears to decrease as rainfall increases.

Y axis

Percentage of
Productivity
Increase



X axis

Inches of Rainfall
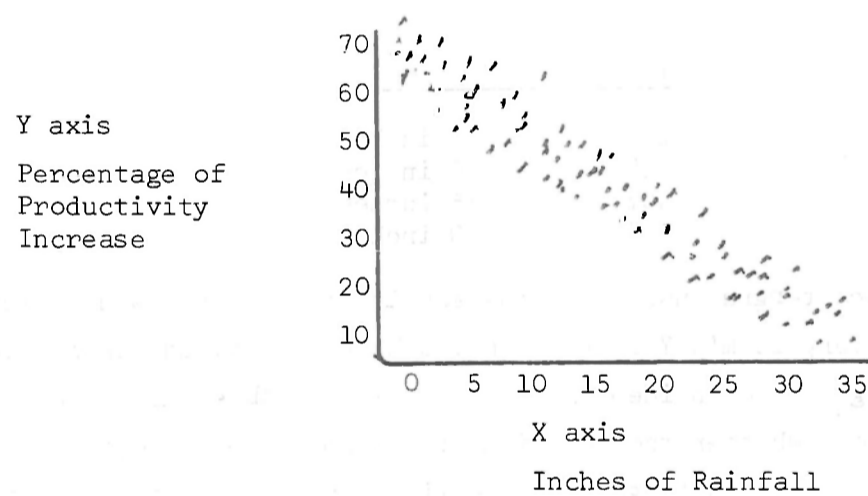
Regression Line: We discuss the regression line in a non-technical way, though
the reader should understand that a regression line has very exact mathematical
properties. It is possible, however, to understand the regression line without
covering these properties in detail. A regression line is a straight line
drawn through a scattergram in such a way that the distance between the line and
all the dots in the scattergram is kept to a minimum.

This last point has to be stated a bit more technically. The regression
line minimises the average of the squared distances between the line and all the
dots in a scattergram. It is for this reason that the regression line is often

called the least-squares line.  And to get the language closer to the standard
vocabulary of statistics, we should speak not of squared distances but of
squared deviations.

We start with a scattergram which plots two different scores or measures
taken on a series of units.  In our example the units are farms and the scores
inches of rainfall and increase in yield after using a hybrid seed.  Above we
plotted the scores of four farms on a scattergram.



Y axis

Percentage of
Productivity
Increase

X axis

Inches of Rainfall
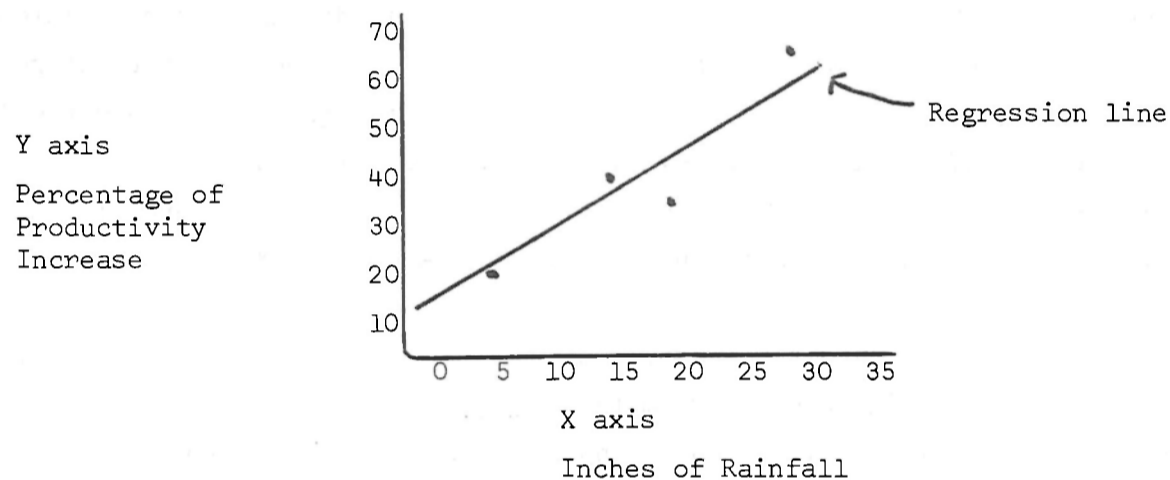
On this scattergram we have drawn an illustrative regression line.
It is a line which minimises **the** distance between each of the four dots and
the line.  A line drawn at any other angle would generate greater average
distances; or, again more correctly, greater average squared deviation.

By now the reader will be wondering what a squared deviation from
a line is, though the concept should not be too difficult if it is remembered
that a standard deviation also starts with the squared deviation of each
score from the mean.

To grasp the squared deviation notion think of each farm as having:
    An X score: the inches of rainfall in the area;

    A Y score: the amount by which the farmer increased maize yield; and
    **A Yc score:**     a new yield score which is a computed score based
                        on the amount of yield predicted from information
                        about rainfall.
The Yc is the new idea here.  The Yc (computed Y score) is the point at which the
X value of a farm intersects with the regression line.

Go back to the previous example, where the scores on four farms were plotted and a regression line drawn. What is the Yc of the farm with five inches of rainfall? You find this out by reading directly up from the score of 5 inches on the X axis to the regression line. When you get to the point where the regression line intersects with an imaginary line drawn directly up from 5 inches, you then read across to the Y axis. It is this point on the Y axis which is the Yc, or the new computed score on Y.

After finding the Yc for each of the farms we would calculate the distance between the Y and the Yc for each farm. This distance is calculated in terms of whatever is measured by the Y axis, in our example by percentage increase in productivity. Thus, for instance, if the regression line directly above 20 inches of rainfall is parallel to 40 per cent increase in productivity, then 40 is the Yc for 20 inches. But the actual Y score of the farm with 20 inches is 35. The distance between Y and Yc is then 5, Y - Yc = 5.

We now see what is meant by least-squares. The deviation of every score in the entire distribution from the regression line is squared. The average squared deviation is a smaller number than would be true were the regression line drawn at any other angle. Conceptually this is similar to the point made previously about the deviation from the mean in a distribution of scores. If you take the deviation of each score in a distribution from the mean of the distribution, square these deviations, **and add them up, the resulting** number will be smaller than the resulting number would be had you taken the deviations from any score other than the mean.

The notion of least-squares will be important in a moment, but first a concluding word on the regression line. The regression line is computed from the actual scores on the two variables, not simply drawn through the scatter-gram by visual estimate. Similarly the Yc would not actually be measured as our example suggests. The Yc would be derived from formulae as applied to the raw data. The appropriate formulae can be found in any statistics text, and are not included here because our goal is conceptual understanding not computational instruction.

The Standard Error (also Known as The Standard Error of Estimate)

We are now ready for a new statistic, the standard error. Each farm has a Y score, its actual yield increase, and a Yc score, its computed yield increase. The computed score is given by the amount of yield increase expected of a farm at a given point on the X axis (rainfall) derived from the regression line.

To understand the standard error we first must understand the idea of dispersion around the regression line. The regression line is in effect a mean score; there is a dispersion around the regression line just as in the original distribution there was dispersion around the mean. The dispersion is simply the sum of the distances between the Y and Yc for each farm in the sample. Consider the following series of steps:

1. A scattergram plots the intersection point for two different scores of each farm in the sample.

2. A line is drawn through that scattergram at an angle which minimizes the averaged (squared) distances of each individual dot from the line.

3. Because not every dot will fall right on the line (unless the correlation is perfect) there will be a dispersion of dots around the line.

4. The standard error is a summary statistic which measures this dispersion. It reveals the amount of error one would make if one predicted the Y score from information about the X score; the error being the distance between the true Y and the computed Y.

The formula which best illustrates a standard error is presented below:

$$\text{Standard error (Sy)} = \sqrt{\frac{(Y-Yc)^2}{N}}$$

You can see that it resembles the formula for the standard deviation. In both cases the difference between a score and a computed score (mean in the standard deviation; regression line and Yc in the standard error) for each unit measured is squared. These squared deviations are summed, divided by the number of units **measured** and then the **square** root is taken.

Not only is the computation of the standard error similar to the standard deviation, its properties are similar.

If the Y scores are normally distributed around the regression line, then 68% of all Y scores will fall within plus or minus one standard error, and 96% will fall within plus or minus two standard errors, just as is the case for the standard deviation in a normal distribution.

The standard error of estimate is measured in the same dimension as the Y axis, as can be seen from the formula. In our example, the standard error would tell us the range of increase in productivity associated with any given amount of rainfall. If, for instance, the standard error were 3, we would know that approximately 68 per cent of the farms with, say, 15 inches of rainfall, would increase their productivity within plus or minus 3 per cent of the Yc, which corresponds to 15 inches.

Let us consider the standard error and its computation from a slightly different perspective. We start with two farms whose X and Y scores are as follows:

|       | X score   | Y score           |
|-------|-----------|-------------------|
| Farm 1 | 10 inches | 40% yield increase |
| Farm 2 | 30 inches | 25% yield increase |

We plot these farms on a scattergram, draw a regression line, and then by use of dotted lines show the deviation for each farm from the regression line.
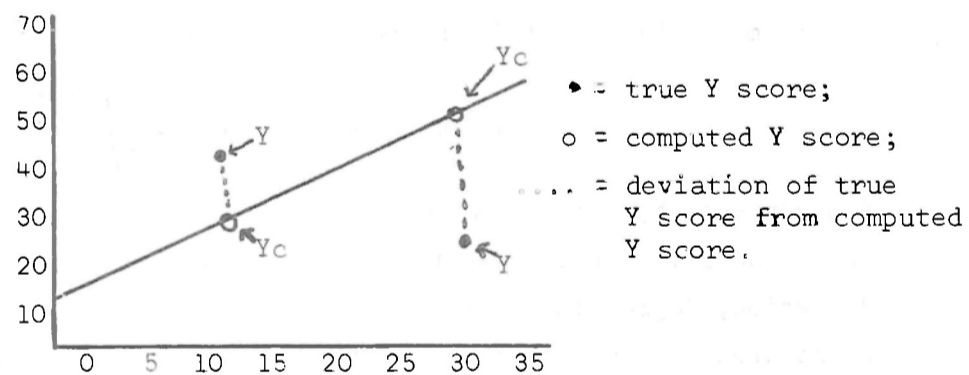


Mathematically, the standard error sums the dotted lines for all 250 farms in the sample. If every farm fell exactly on the regression line there would be no dotted lines to sum, and thus the computational formula would produce a 0 for the standard error. If the farms are very dispersed around the regression line, the sum of the dotted lines is large and the standard error is large as well.

Recall now the idea of a normal distribution around the regression line. We said that approximately two-thirds of the cases will fall within one standard error. A low standard error (just as a low standard deviation) would indicate a great deal of clustering around the regression line. A high standard error would indicate a great deal of dispersion around the regression line.

Here are two illustrations, one showing the standard error to be 3 per cent; the other showing the standard error to be 10 per cent. (Remember that the standard error is a measure of whatever is measured by the Y axis, in our **examp**le by the percent increase in yield.)

Standard error = 3%



+ 1 Sy
- 1 Sy

Regression Line

Standard error = 10%



+ 1 Sy
- 1 Sy
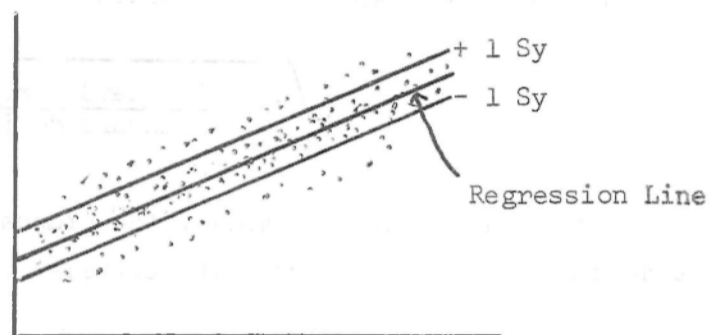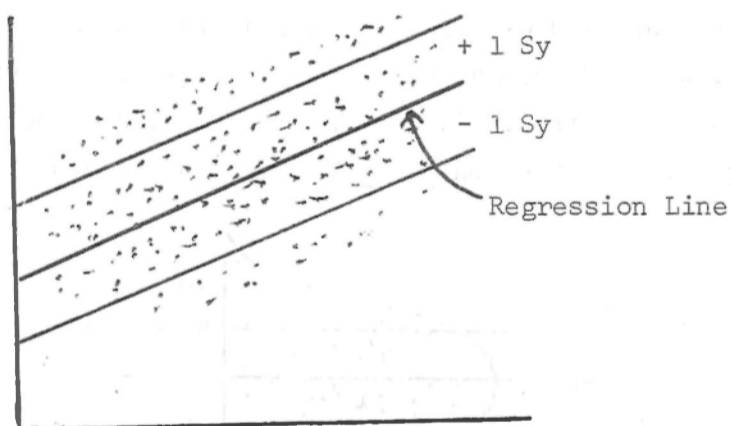
Regression Line

In both diagrams approximately two-thirds of the farms fall within plus or minus one standard error of the regression line. In the top diagram, however, the farms are more tightly clustered around the regression line than they are in the bottom diagram (which is why there is a lower standard error in the top than in the bottom diagram). This has a very practical consequence. In predicting the yield of a farm from information about its rainfall, the size of the error would on the average be much smaller if the data are represented by the top rather than the bottom diagram. Stated differently, the lower the standard error the more one knows about the dependent variable based on information about the independent variable.

Correlation Coefficient

       The correlation coefficient (often indicated by the letter r) is
a statistic indicating the direction and the strength of a relationship
between two variables.  This statistic is really rather simple once deviation
and the standard error is understood, for the correlation coefficient is nothing
more than the ratio between these two measures.  Here is the formula:

$$r = \sqrt{1 - \frac{\text{Standard error squared}}{\text{Standard deviation squared}}}$$

       You can see that the smaller the standard error is in relation to
the standard deviation, the higher the correlation.

       The way to understand this is to think all the way back to our first
distribution on the Y axis, that is, the variability in yield production.  When
we thought only in one-dimensional space (that is only about one variable),
the dispersion we were interested in was the dispersion around the mean.  The
measure we used was the standard deviation, which was 12 per cent.  And the
mean, we said, was 40 per cent.  On the average yield was improved by 40 per
cent after introduction of the hybrid seed.  The following diagram shows the
original distribution before being plotted against the second variable.
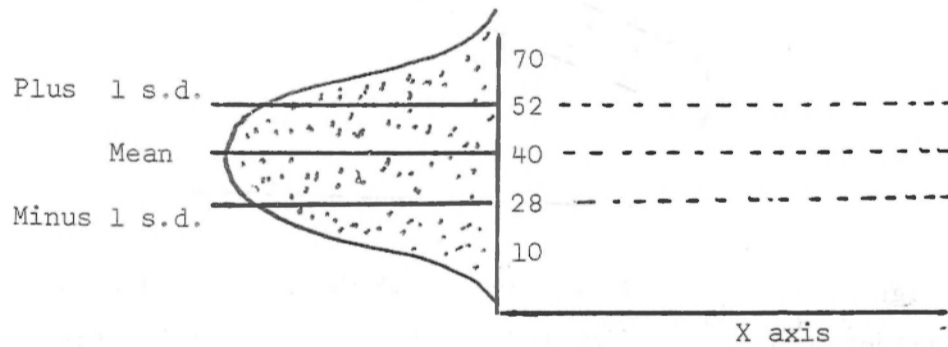


       This diagram reminds us that the original scores were distributed
around the mean.  But when we computed the standard error we learned that the
scores also were distributed around the regression line.  If the scores
cluster more tightly around the regression line than they do around the mean
of the original distribution, then there will be a correlation between the
X and Y variables.  Study carefully the following diagram.

Y variable

+1 standard error
-1 standard error

+1
s.d.

Mean of Y

-1
s.d.

X variable
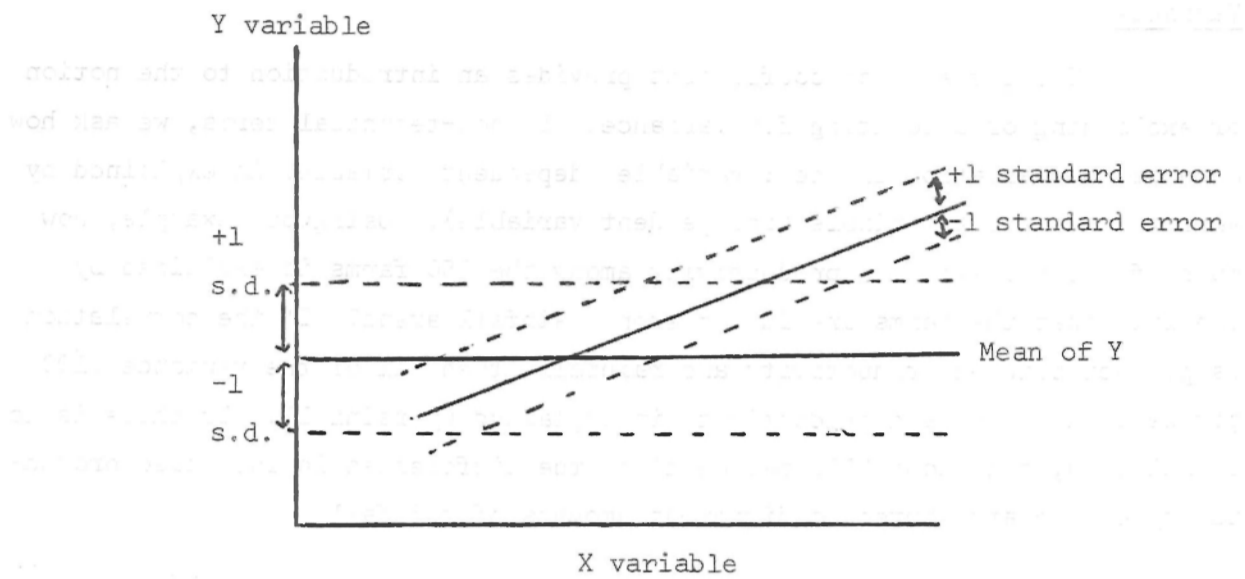
Here we see that the standard deviation is much larger than the
standard error; stated differently, the scatter of scores around the mean of
the original distribution is much larger than the scatter of scores around
the regression line. There will be a high correlation between X and Y.

Now compare the following diagram.

Y variable

+ 1 Sy

+ 1
s.d.

- 1 Sy

Mean of Y

- 1
s.d.
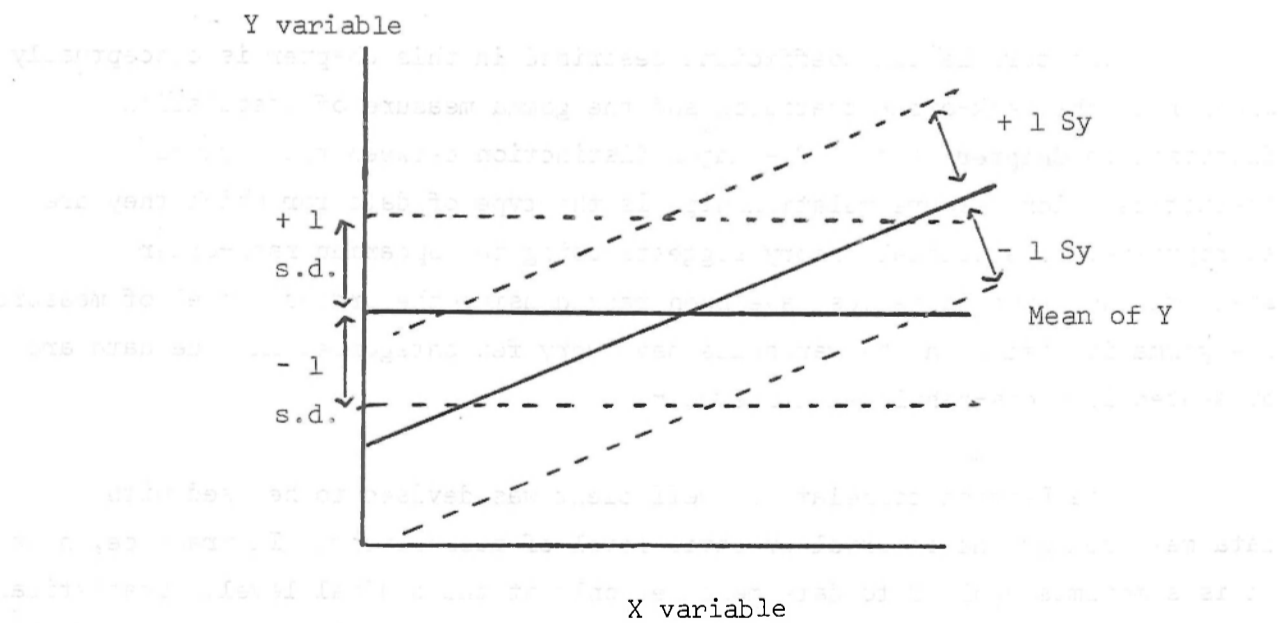
X variable

Here the standard deviation and the standard error are more nearly
the same in size, and thus the correlation will be lower. Put differently, we
do not learn too much about the original distribution of the Y variable by
using X as an independent variable.

<u>Variance</u>

The correlation coefficient provides an introduction to the notion of explaining or accounting for variance. In non-technical terms, we ask how much of the variation in the Y variable (dependent variable) is explained by variation in the X variable (independent variable). Using our example, how much of the variation in productivity among the 250 farms is explained by the fact that the farms are in different rainfall areas? If the correlation is perfect between productivity and rainfall, then all of the variance (100 per cent) in increased productivity is explained by rainfall. If there is no correlation, then none ( 0 per cent) of the differences in increased productivity can be attributed to different amounts of rainfall.

The percentage of variance explained is given by the square of the correlation coefficient. If X and Y are correlated at .3, then 9 per cent of the variance in Y is explained by X. **If they are correlated at .7, then 49 per cent of the variance is explained, and so forth. As can be seen, it takes a high correlation to explain much of the variance.**

<u>Conclusion</u>

The correlation coefficient described in this chapter is conceptually similar to the rank-order statistic and the gamma measure of association discussed in **Chapter** Seven. The major distinction between the several statistics which measure relationships is the type of data for which they are appropriate. Statistical theory suggests using the Spearman rank-order statistic when two variables have been ranked using the ordinal level of measure. The gamma is used when the variables have very few categories and the data are presented in cross-tabulated table form.

The Pearson correlation coefficient was devised to be used with data measured at the interval or ratio level of measurement. In practice, however, it is sometimes applied to data measured only at the ordinal level. Statistical theorists advise against this practice; but some social scientists report that the results are approximately the same whether one uses a rank-order statistic or the Pearson coefficient on ordinal data.

The discussion of the appropriate application of various statistics is a complicated one and is beyond the scope of this text. Our purpose has been to help the reader understand when a researcher reports that 'the correlation between social status of citizens and their level of political participation is .64' or that 'there is a negative correlation at the .47 level between mechanisation of farms and distance of farms from urban centres'.

The Pearson coefficient is not only widely used in studies which make such assertions, it is a statistic with many more advanced applications. Partial correlation analysis or multiple regression analysis, for instance, are advanced techniques based on simple correlation analysis of the kind reviewed in this chapter.

At the conclusion of the review of the gamma measure of association it was suggested that the reader try to visualise the type of percentage table on which a given gamma measure might have been computed, even when the raw data are not presented.

The same suggestion holds for understanding and interpreting correlation coefficients. Except now it is necessary to visualise a **scattergram** rather than a cross-tabulated table. Here is a guide which will help in this exercise:

1. Identify the two variables which are being correlated.

2. Picture these two variables as a scattergram, the independent variable along the horizontal axis and the dependent variable along the vertical axis.

3. If the correlation is positive, picture the scattergram as showing how a higher score on one variable is associated with a higher score on the other variable.

   If the correlation is negative, picture the scattergram as showing how a higher score on one variable is associated with a lower score on the other variable.

4. Depending on whether the correlation is weak or strong, picture the dots on the scattergram as either widely scattered around the **regression line or as tightly clustered around it.**

When the researcher reports that there is no correlation between two variables, picture a scattergram in which the dots show no pattern at all.

A bit **of** practice at this game will make anyone a more efficient and intelligent reader of data analysis using correlational statistics.

CHAPTER 10

STATISTICAL SIGNIFICANCE

The statistics thus far discussed have primarily been measures of association or correlation. They are used to determine and then summarise the kind of relationship we find between two variables.

Statistics perform a second very different task for the researcher. Statistics can tell the researcher how much confidence to place in his findings. We have already had an introduction to this use of statistics in connection with sampling theory. At the conclusion of Chapter Two there is a discussion of sample error and confidence intervals, and toward the end of Chapter Eight there is a discussion of confidence intervals in connection with the idea of a normal distribution.

Both the discussion in Chapter Two and in Chapter Eight are part of a larger statistical topic often summarised as statistical inference. In this chapter we take up a further idea associated with statistical inference, the idea of statistical significance. All of these discussions are closely interrelated, and all of them derive from some basic ideas in sampling theory.

## Statistical Significance

As we turn to one of the most complicated issues in social statistics we immediately encounter a terminological confusion. This confusion surrounds the term significance. We all know the dictionary meaning of significance. It is used synonymously with the term importance, as when we say that the Arusha Declaration was significant in the political history of Tanzania.

The usage which equates significance with importance plays a role in commonsense social science. We refer to a significant finding when the finding is thought to be important for scientific understanding or for policy reasons. Thus it would be significant to learn that a fertiliser increases maize yield by 50 per cent, significant because we value increased yield as a means of rural development. This is to use the term significance in the substantive sense, as when we are concerned about the substantive findings **of social research.**

The term significance has a different meaning in statistical usage. Its meaning in statistics is quite technical and not necessarily consistent with commonsense usage. For instance, a very trivial finding in social science may nevertheless be statistically significant. And a finding which impresses us as socially important may be reported as not statistically significant.

A bit of effort can clear up the confusion.  Statistical tools
do very specific things, as we have already seen.  One of these things is
to tell the researcher whether a particular finding could have been an accident.
It is in this context that the term statistical significance is used.  The term
has a precise meaning which, when understood, greatly increases one's ability
to read and comprehend scientific reports which use statistics.

The meaning of statistical significance, briefly summarised,
refers to the likelihood that a particular finding could have happened by
chance.  The statistician is a skeptic.  He assumes that any empirical finding
is a chance finding until it can be demonstrated otherwise.  He asks, is the
result reported a <u>significant</u> departure from chance expectation?  The test he
sets up presumes that what is observed is just a chance happening.  He then
measures the extent to which the results reported deviate from chance.  If they
deviate in a substantial manner, he concludes that the result is statistically
significant.

Try a simple illustration.  Eighty farmers are included in a study
of hybrid seed.  Half of these farmers are asked to plant with the new seed;
the remaining half do not use the new seed.  One year later the research team
finds out that the yield of 50 per cent of the farmers has increased while
the yield of the rest of the farmers has not changed.  The question
immediately comes up, did the farmers using the hybrid seed increase their
yield at a greater rate than those not using the seed?

If we found Table A to be the case, we certainly would conclude
that the hybrid seed increases the yield:

Table A

|  | Farmers Not Using the Hybrid Seed | Farmers Using the Hybrid Seed |
|---|---|---|
| No Yield Increase | 40 | 0 |
| Yield Increase | 0 | 40 |
| N= | (40) | (40) |

Every farmer who used the seed ended up with a yield increase; every
farmer not using the seed failed to increase the yield.  We would not need
a complicated statistical test to tell us that the seed had an effect.

If the results were as shown in Table B the conclusion would again
be unambiguous.  There is no relationship between the use of the seed and whether
a farmer increased his yield.  We might say that there is a completely random

Table B

|  | Farmers Not Using the Hybrid Seed | Farmers Using the Hybrid Seed |
|---|---|---|
| No Yield Increase | 20 | 20 |
| Yield Increase | 20 | 20 |
| N= | (40) | (40) |

relationship between the use of hybrid seed and the increase of yield.

Thus far there is no difficulty in drawing conclusions and no reason to introduce statistical tests. What, however, if after the year the results of the study show what appears in Table C?

Table C

|  | Farmers Not Using the Hybrid Seed | Farmers Using the Hybrid Seed |
|---|---|---|
| No Yield Increase | 25 | 15 |
| Yield Increase | 15 | 25 |
| N= | (40) | (40) |

Now we learn that among the farmers using the new seed, more increased their yield than failed to increase it. And the opposite is true of those not using the seed. But is the finding convincing? After all, as many as 15 farmers who used the seed failed to improve yield. And among those not given the seed there was a substantial minority (37%) who nevertheless did increase yield. Perhaps the hybrid seed really makes no difference. Perhaps it was just an accident that a few more farmers using the seed than not using the seed managed to increase yield.

Perhaps the finding is just a chance result. This is what the skeptical statistician would say. He would continue to say this until it had been demonstrated that the result observed (as shown in Table C) significantly departs from chance. The statistical tools used to demonstrate this are tests for statistical significance. They tell us the likelihood that a finding such as reported in Table C is something other than accident.

The Chi-Square Statistic

Pronounced "ki-square" and symbolised as $X^2$ (the greek letter chi), the chi-square is a very frequently used statistic in social research. It is an excellent statistic for further explaining the notion of statistical significance, though as a teaching tool it has a few drawbacks.

One difficulty in using the chi-square statistic is that it draws on
some language which goes against commonsense. And it is never easy to understand
a new idea which must be explained in language used in an unorthodox and even
counter-intuitive way. So from the beginning we will stress that the language is
more confusing that the statistic. If we can penetrate the language, we will
quickly understand the statistic. And as soon as we understand the statistic we
will have grasped the idea of statistical significance.

The chi-square measures whether something observed differs significantly
from something expected.

The initial language difficulty is in the term expected, for in the
above sentence it is not used in a commonsense way. Mostly we set out to do
research because we expect some findings. For instance, if we were studying
the introduction of a hybrid seed we would expect there to be a relationship
between its use and increased yield. Unfortunately, the term expected in
technical statistical language does not have this commonsense meaning, but
rather the opposite.

The term expected in the definition of the chi-square statistic means
that we assume no relationship. It is as if we said, "We will check to see
if the farmers using the hybrid seed increased yield, but we don't really expect
that they did. We doubt if there is any relationship between the seed used
and yield." In technical terms such a sentence would be stating the null
hypothesis, a somewhat confusing phrase but so widespread in social research
that we cannot avoid it. The null hypothesis is the assumption that there is no
relationship between the variables one is studying.

In tests of statistical significance we actually try to disprove
the null hypothesis. If we can disprove it, then we conclude that there is a
relationship.

Let us use once more the example of the hybrid seed and yield
increase. First, I expect (or hypothesise) there to be no relationship between
the use of the hybrid seed and yield increase. Stated differently, I expect to
find what Table B shows above, for it perfectly expresses the null hypothesis.
Second, I observe the actual relationship between use of the hybrid seed and
increase in yield. Third, I measure the extent to which **the** observed results
differ from the expected results, keeping in mind that I expect to find nothing.
This is what the chi-square statistic measures. Fourth, if the observed results
and the expected results are similar, I conclude that introducing the hybrid

seed does not increase yield. In short, I accept the null hypothesis of no relation-
ship. If, however, the observed pattern significantly differs from the expected
result of no relationship, then I conclude that the hybrid seed does increase
yield. In technical language, I reject the null hypothesis and thereby
accept that there is a relationship between the two variables.

Just how different a result has to be to be significantly different
is something we take up after looking more closely at the workings of the chi-
square statistic.

## Computing a Chi-Square

The actual computation of the statistic involves calculating the
difference between no relationship, called the expected result, and the
relationship one finds, the observed result. In the formula shown below, the
0 stands for the observed result and the E stands for the expected result.

$$x^2 = \frac{(0-E)^2}{E}$$

where E = expected frequency
        0 = observed frequency

The actual computation is simple. Let us assume that Table C
reported above is what we actually observe to be the relationship between
hybrid seed usage and yield increase. We can compare this Table with Table B,
which represents the null hypothesis or no relationship.

Each Table has four separate groups of farmers, and for each group
we measure the deviation between expected and observed results. These deviations
are squared, divided by the expected result, and then summed to get the chi-
square. The figures below show the computation:

| Farmers Hybrid Seed | Yield Increase | Observed Frequency (0) | Expected Frequency (E) | Difference (0-E) | Difference Squared $(0-E)^2$ | Diff. Squared Divided by Exp. Freq. $(0-E)^2/E$ |
|---|---|---|---|---|---|---|
| NO | NO | 25 | 20 | 5 | 25 | 1.25 |
| YES | NO | 15 | 20 | -5 | 25 | 1.25 |
| NO | YES | 15 | 20 | -5 | 25 | 1.25 |
| YES | YES | 25 | 20 | 5 | 25 | 1.25 |
| | | | | | | $x^2$= 5.00 |

The chi-square for Table C is 5.00, a number which as yet has no
meaning to us. In order to make use of this number to determine whether Table C
is statistically significant we must introduce a few more ideas.

The first of these ideas is degrees of freedom, often symbolised
simply as df. We will not attempt a technical discussion of this concept.
Knowing how to determine the degrees of freedom for any table is sufficient
for the present purposes. If you were presented the following table, could
you fill in the remaining numbers? Using technical language, could you fill

|  | Farmers Not Using<br>The Hybrid Seed | Farmers Using the<br>Hybrid Seed |  |
|---|---|---|---|
| No Yield Increase | 25 | ? | N =<br>40 |
| Yield Increase | ? | ? | 40 |
|  | N =   40 | 40 | 80 (total) |

in the cell frequencies from knowledge you have of the marginal frequencies.
The marginal frequencies appear at the margins of the table; they give
the total number of cases having any given characteristic. The marginal
frequency for farmers not using the hybrid seed is 40, and the marginal frequency
for farmers using the hybrid seed is also 40. The table also gives the marginal
frequencies for the two categories of no yield increase and yield increase.
Inside the table are the cell frequencies. In a table which cross-tabulates
two variables each with two categories there will be four cell frequencies.

The above table shows only one cell frequency, but it is obvious
what the remaining three cell frequencies are. For instance, the number of
farmers not using the seed and yet increasing yield must be 15, because 25 + 15 = 40.
And so on. All the cell frequencies save one (in a 2 by 2 table) are determined
as soon as one cell frequency is inserted.

Hence the idea of degrees of freedom. In a table such as Table C,
there is only one degree of freedom, because only one cell frequency is free
to take any value. As soon as that cell takes a value, any value, all
remaining cell frequencies are determined. It is relatively easy to determine
the degrees of freedom in a four-cell table, but what about tables where the
variables have more categories? For instance, what if there were a third
group of farmers, those who used a mixture of hybrid and non-hybrid seed, and
yield were measured in terms of four categories: no yield, low yield, moderate
yield, high yield. The table would then look like this:

|  | Farmers Not Using<br>the Hybrid Seed | Farmers Using a<br>Mixture of Seeds | Farmers Using<br>Only Hybrid Seed |
|---|---|---|---|
| No Yield |  |  |  |
| Low Yield |  |  |  |
| Moderate Yield |  |  |  |
| High Yield |  |  |  |

This table has 3 times 4, or 12 cells.   It would be cumbersome to calculate
the degrees of freedom through a trial and error method of actually filling
in cell frequencies until all the remaining were determined.

Fortunately there is a very simple formula for calculating degrees
of freedom.   It follows the convention of calling the categories across the
top of the table the column categories (C) and calling the categories down
the side of the table the row categories (R).   The formula is:

degrees of freedom = (C-1) X (R-1)

Or, the number of columns minus one times the number of rows minus one.  Using
the above table we would get (3-1) X (4-1) = 6 degrees of freedom.   That is,
once six cell frequencies had been entered into the table, all remaining cell
frequencies would be fixed.   The same formula also applies to the 2 by 2 table
we considered first:

(2 - 1 ) X (2 - 1 ) = 1.

This detour into an explanation of degrees of freedom has been
necessary if we are to make sense of the chi-square computed above.   It will be
recalled that Table C gave us a chi-square of 5.00; and we said that this number
could tell us whether the data presented in that table significantly differed
from chance expectation (the expectation that there is no relationship between use
of hybrid seed and **yield** increase).   The initial step in making sense of the number
5.00 was to compute the degrees of freedom, which we found to be one for this
table.

The next step is to turn to something called the $x^2$ distribution.
But before actually examining this distribution, we think a bit about what we
expect to find.   Another detour is necessary, this one back into the concepts of
sampling theory.

First, we start with the relationship between two variables: the use of
hybrid seed and yield increase.

Second, these two variables can be related in **many** different ways;
three of these ways are shown in Tables A, B  and C.   But of course other
illustrative tables could be easily created.

Third, for every possible way in which the two variables can be
related there is a corresponding $x^2$.   That is, we computed a chi-square only for
Table C, but could just as easily compute one for Tables A and B, or any other

table which could be derived from the two variables presented.

Fourth, statisticians have worked out the probability of getting any particular chi-square value.

It is this fourth point which is most important. Statisticians tell us that there is a population of $X^2$ values, or more correctly a population of $X^2$ values for any given degree of freedom. For instance, for tables with one degree of freedom there can be a chi-square value anywhere from 0.0 to 10.8 (to state a given upper limit is not exactly correct, but it will serve for present purposes). All numbers between 0.0 and 10.8 constitute the population of $X^2$ values for tables with one degree of freedom.

But of course in research we are only dealing with one table at a time. The idea of a population of $X^2$ values may be of importance to the statistical theorist, but what does it tell us about our particular finding? There is a simple answer to this question. Statistics treats our particular finding, and its $X^2$ value, as a sample from the population of $X^2$ values. And statistics can tell us the probability of getting a $X^2$ value as high or higher than we found.

Le us look at a modified distribution of $X^2$ values to see how this works. A table showing the distribution of $X^2$ has three major parts: first, there is a column labeled df or degrees of freedom; second, there are a series of column headings which give different probabilities, such as .30 or .05 or .001; third, there are the $X^2$ values themselves, with a particular $X^2$ being associated with each degree of freedom and probability.

Modified $X^2$ Distribution

| df | .50 | .30 | .20 | .10 | .05 | .02 | .01 | .001 |
|----|------|------|------|------|------|------|-------|-------|
| 1  | .46  | 1.07 | 1.64 | 2.71 | 3.84 | 5.41 | 6.64  | 10.83 |
| 2  | 1.4  | 2.41 | 3.22 | 4.60 | 5.99 | 7.82 | 9.21  | 13.82 |
| 3  | 2.4  | 3.66 | 4.64 | 6.25 | 7.82 | 9.84 | 11.34 | 16.27 |

etc.

Across the top of this table are listed probabilities. A probability of .50, for instance, means that the chances are exactly 50 per cent that a given result would occur. A probability of .05 means that the chances are only 5 in 100, or 5 per cent, than a given result would occur. And a probability of .001 means that the chances are only 1 in 1,000 that a given result would occur.

Now let us see how to use the $X^2$ distribution. First we go to the row of figures associated with the df of our table. Because our table had

only one df, we are at present interested only in the top row of the
distribution. We next look across this row for the $X^2$ value which is equal to
or larger than the $X^2$ computed. The $X^2$ computed for Table C was 5.00, and thus
we look across until we find a 5.00 or figure higher than 5.00. We quickly see
that a value of 5.00 falls somewhere between the 3.84 associated with the .05
level of probability and the 5.41 associated with the .02 level of probability.

It is the convention always to work with the level of probability
associated with the $X^2$ just lower to the one computed (unless the computed
$X^2$ happens to correspond to an exact $X^2$ value in the table). For our purposes,
then, we say that the probability of getting a $X^2$ as high or higher than 5.0
is .05. Stated differently, the relationship reported in Table C is statistically
significant at the .05 level.

We are now ready to take up again the idea of statistical significance,
which is where we began our discussion of the $X^2$. A finding is statistically
significant when it departs from chance. If the finding does not depart from
chance, we accept the null hypothesis that there is no relationship between
the variables.

In the example we have been using, chance is shown in Table B, where
the farmers using the hybrid seed showed no more and no less yield increase
than the farmers not using the seed. Another name for chance is the expected
finding if there is no relationship. In our examples the chances of a farmer
increasing his yield are not at all affected by whether he uses the hybrid
seed.

We carry out the research, and find what is reported in Table C.
But the skeptic persists: "You haven't really found out that the hybrid seed
increases yield. The differences in Table C are not all that big. Probably
it was just an accident that a few more farmers using the hybrid seed happened
to increase yield."

The $X^2$ statistic helps us answer the skeptic. It tells us that the
findings in Table C do depart from chance; the observed frequencies differ from
the expected (chance) frequencies. But how substantially do they depart from
chance?

Stated precisely, we are 95 per cent certain that Table C is a
true finding and that the use of the hybrid seed does increase productivity.
Yet we are 5 per cent uncertain. Five per cent of the time a $X^2$ as large

as the one we found (in tables with one degree of freedom) would be an accident.
If we had found a larger $x^2$, say 8.5, then we would be more confident that the
hybrid seed makes a difference. For a $x^2$ of 8.5 would happen only 1 in 100 times
by accident. And a $x^2$ larger than 10.83 would be better yet, for we then would
conclude that our finding could happen only 1 in 1,000 times by accident.

But we are stuck with what we found, a $x^2$ value which could happen
by chance 5 out of 100 times. Whether this is enough certainty to distribute
hybrid seed throughout the country is of course not a statistical question.
It is a question of politics and economics and judgment. Statistics can tell
us how likely it is that a relationship between variables is a true relationship.
Statistics cannot tell us how to act once we have this information. If the
hybrid seed were enormously expensive, being confident of our finding at the
.05 level may be insufficient. Even though there is only a 5 per cent chance
that the hybrid seed does not increase yield, the costs of being wrong are too
great.

Note, however, that had the $x^2$ been 10.83 uncertainty would have been
reduced but not eliminated. The finding could still have happened by chance one
time out of one thousand. If our particular sample happened to be that one time,
we would be wrong in concluding that the hybrid seed increases yield.

A Further Illustration of the Chi-Square

The actual example used to illustrate the computation and interpretation
of a chi-square used a very simple 2 by 2 table. Some readers might wonder how
a chi-square is used when data are more complicated. In fact, the computation
requires more arithmetic but is not much more complicated than what was
illustrated by the simple table. And the interpretation of the chi-square is
identical.

In a study of 786 African students in Nairobi secondary schools
(in 1967) the following question was asked:

Which sentence is more true?
1. Ordinary people should feel free to give advice to their
   **political** leaders, or to ask them for help.

2. Political leaders cannot do their work properly if ordinary
   people are always giving them advice or asking for help.

We found that 79% choose the first sentence and 21% choose the second.  We
wanted to learn whether students in different Forms answered differently.
Here were the results:

|            | Form of Students |      |     |     |
|            | II   | IV    | VI  |     |
|------------|------|-------|-----|-----|
| Sentence 1 | 164  | 421   | 37  | 622 |
| "        2 | 60   | 99    | 5   | 164 |
|            | 224  | 520   | 42  | 786 |

Earlier we distinguished between marginal totals and cell frequencies.
The marginals describe the population in terms of the total number of persons
taking any particular value on a variable.  For example, one variables is "form
of student" and this variable has three values (or categories):  Form II, Form IV
and Form VI.  One set of marginals then is the number of students in each Form --
224, 520  and 42 respectively.  Another set of marginals is the number of respondents
choosing Sentence 1 (622) and the number choosing Sentence 2 (164).  The marginals
are at the margins of the table.  The two sets of marginals will total to N (786
above).

Cell frequencies are the results obtained from the cross-tabulation.
There is a separate cell frequency for each value of the independent variable as
it intersects with each value of the dependent variable.  Thus there are 164
Form II students who choose Sentence 1.  This is a cell frequency, together with
all the remaining numbers in the cells within the table.  Since this is a 3 by
2 table, there will be six cell frequencies.

Chi-square tells us how much an observed relationship differs from
an expected one; we state the expected relationship as follows:  "We expect
there to be no relationship between the two variables."  (To repeat, this is the
null hypothesis.)  Stated differently, because we know that 79 per cent of the
total population chooses Sentence 1, the null hypothesis (of no relationship)
says that 79 per cent of each Form will choose that sentence.  The proportion in
the cells will look exactly the same as the proportion at the margins.  Here is
the table with the individual cells designated by small letters and individual
marginals designated by capital letters.

|            | Form of Student |            |           |                 |
|            | II              | IV         | VI        |                 |
|------------|-----------------|------------|-----------|-----------------|
| Sentence 1 | $a$ 164         | $b$ 421    | $c$ 37    | 622(A)          |
| Sentence 2 | $d$ 60          | $e$ 99     | $f$ 5     | 164(B)          |
|            | 224(C)          | 520(D)     | 42(E)     | 786(N, or number) |

The first step then is to determine what would be the frequency
in each cell if there were no relationship (called the expected frequency).
There are several procedures for doing this, the most common being:

Multiply the two marginals which intersect for any given cell and
divide by N, the total number.

Doing this would tell us, for instance, how many Form II students
would choose Sentence 1 if there were no relationship between Form of student
and choice between the sentences. Stated differently, how many students would
be in cell a if the relationship between the two variables is random? The number
in cell a under such conditions is the expected frequency.
Thus,

$$
\begin{array}{llllll}
\text{Expected frequency for cell} & a & = (622\times224)/786 & = & 177 \\
" & " & " & " & b & = (622\times520)/786 & = & 411 \\
" & " & " & " & c & = (622\times 42)/786 & = & 33 \\
" & " & " & " & d & = (164\times224)/786 & = & 47 \\
" & " & " & " & e & = (164\times520)/786 & = & 109 \\
" & " & " & " & f & = (164\times 42)/786 & = & 9 \\
\end{array}
$$

These figures can be checked easily enough. Because we know that
79 per cent of the total population of respondents choose Sentence 1, the same
proportion in each Form should do so if there is no relationship. Thus the null
hypothesis tells us that 177 students from Form II will choose sentence 1;
177/224 = 79 per cent, etc.

The formula for the chi-square **calls** for squaring the difference
between the observed and expected frequencies, and dividing these differences by
the expected frequency. The sum of these will be the chi-square. We set the
computation up as follows:

| Cell | Observed Frequency (O) | Expected Frequency (E) | Differ-ence (O-E) | Difference Squared $(O-E)^2$ | Difference Squared divided by Expected Freq. $(O-E)^2 / E$ |
|---|---|---|---|---|---|
| a | 164 | 177 | -13 | 169 | .955 |
| b | 421 | 411 | 10 | 100 | .243 |
| c | 37 | 33 | 4 | 16 | .485 |
| d | 60 | 47 | 13 | 169 | 3.596 |
| e | 99 | 109 | -10 | 100 | .917 |
| f | 5 | 9 | -4 | 16 | 1.778 |
| total | 786 | 786 | 0 | | sum 7.974 |

For this table then, $X^2 = 7.974$.

In order to interpret the chi-square of 7.974 we calculate the degrees of freedom associated with the table. The formula calls for multiplying the rows minus one times the colums minus one; or, (2-1) x (3-1) = 2. This table has two degrees of freedom. According to the abbreviated chi-square table presented earlier for tables with two degrees of freedom a chi-square of 7.97 falls between the chi-square associated with the .02 level of probability and the .01 level of probability.

Following convention, we use the lower level of probability and conclude that the deviation from the expected no relationship is statistically significant at the .02 level. That is, we are 98 per cent certain that the pattern in the initial table did not occur by chance.

## A Cautionary Footnote

Statistical significance, as measured by chi-square or any other measure, is not the same thing as statistical correlation or relationship. A very weak relationship can be statistically significant. A test of statistical significance tells us whether a finding could have happened by chance; it does not tell us that the finding is a positive one or negative one, a weak one or a strong one.

Here are two tables showing a relationship between education (dichotomised as low and high) and income (also dichotomised as low and high). In both tables the direction and the strength of the relationship is identical, a gamma of .12. The first table is based on a sample of 200; the second on a sample of 2,000.

|  | Education | |  |  | Education | |
|---|---|---|---|---|---|---|
| Income | Low | High |  | Income | Low | High |
| Low | 53% | 47% |  | Low | 53% | 47% |
| High | 47 | 53 |  | High | 47 | 53 |
|  | 100% | 100% |  |  | 100% | 100% |
| N= | (100) | (100) |  | N = | (1,000) | (1,000) |
| Gamma = .12 |  |  |  | Gamma = | .12 |  |

Although the relationship is identical in the two tables, the likelihood that the finding could have happened by chance is not at all identical. For the first table, the chi-square is .72, not even significant at the .30 level. For the second table, the chi-square is 7.2, significant at above the .01 level.

That is, using the first table we are not very confident that there
is a relationship between education and higher income.  The pattern discovered
could have happened by chance 30 times out of 100.  Yet using the second table
our confidence in the finding goes way up.  In fewer than one time out of 100
could these observed results have happened by chance.

Why does changing the sample size so dramtically increase the
statistical significance of a finding?  After all, the strength of the relationship
itself (as measured by gamma) does not alter at all.

A mathematical explanation cannot be attempted here, but a common-
sense observation will shed some light.  The expected frequency for any cell
in the first table is 50; the observed frequency in each instance differs from
the expected frequency by three.  How likely is it that twelve respondents in
a sample could be misplaced?  That is, how likely is it that some or all of the
three persons presently classified as low on education and low on income really
belong in the high income classification?  At any one of several steps in the
research process it would have been easy to make a mistake: a poor interviewer,
a misunderstood question, a clerical error in recording or coding, an error
in tabulation.  The chances of a mistake are really rather high.  And if there
is a mistake which has incorrectly classified a few respondents, then indeed
the finding that income goes up with education is an accidental finding.  We
do not put much confidence in a finding which could be altered just by shifting
a few persons around.

In contrast, consider the second table, based on a sample of 2,000.
The expected frequency in each cell is 500; and for each cell the number of
respondents who differ from the expected frequency is 30, or a total of 120
persons.  It is much less likely that 120 respondents have been  incorrectly
classified than that 12 respondents have been incorrectly classified.  Thus
we are more confident that the differences observed represent true differences.

And this is what the chi-square tells us.  We can be fairly certain
(99 times out of 100) that there is a relationship between education and income
if the second table is presented to us.  The relationship itself is not strong.
But we are confident it is a true one.

The logic can be extended.  In the following table, based on a sample
of 20,000, the relationship between education and income is practically

negligible, a gamma of .02, just slightly above zero.  But the chi-square for this
table is 10.00, significant at nearly the .001 level.

| Income | Education Low | High |
|--------|------|------|
| Low | 51% | 50% |
| High | 49 | 50 |
| | 100% | 100% |

Gamma = .02   .  N= (10,000)   (10,000)

Chi-square = 10.00

The reason that such a negligible relationship can still be statistically
significant is again because of the absolute number of people who differ from
the expected pattern.  Altogether 200 respondents demonstrate observed traits
which differ from the expectation that there is no relationship between education
and income. **Two hundred is a** sufficiently large number that it is not likely these
observed results could have happened by chance.

We can summarise the discussion with two points.  First, statistical
significance and statistical relationships are not the same thing.  To be an
intelligent user of social science it is necessary to know the difference.  A
common mistake is to presume that because a finding is said to be **highly statistically**
significant the conclusion is drawn that the relationship must certainly be a
strong one.  Hopefully the present discussion will help you guard against this
mistake.

Second, in Chapter Two there is a discussion of sample size and
confidence in sample estimates.  The assertation is made that the larger the
sample size the higher the confidence you can place in sample estimates.  The
logic of this assertion is analogous to the argument just presented in
connection with sample size and the chi-square.

Comparison of Means

Our lengthy discussion of the $X^2$ statistic introduced several new
ideas: statistical significance, the null hypothesis, deviation from a chance
happening, degrees of freedom, a probability distribution of statistical values.
These ideas are important even if one never uses the $X^2$ because they are
associated with many other statistics used by researchers to determine whether
a particular finding could have happened by chance.  In this section we turn

to another of these statistics which has many uses in social research.

The statistic is the comparison of means.  It often is symbolised as t, because a statistic called Student's t is computed when means are compared.

A research report tells us that the average exam score in a mathematics paper for boys is 56 and for girls is 53.  Are we to conclude from this that boys are mathematically more adept than girls?  An average difference of 3 points is not much; perhaps the difference is due to chance and does not really reflect different mathematical aptitude between boys and girls.

We are told that agriculture extension agents who have only a C.P.E. give an average of 59 per cent of their visits to progressive farmers, while better educated agents give an average of 53 per cent.  This contradicts our research hypothesis, for one might imagine that the better educated agents would spend more not less time than less educated agents with progressive farmers.  Perhaps, however, this is just a chance finding; a difference of 6 per cent between the two groups of agricultural agents is not very great.

The t statistic is designed to tell the researcher whether a difference between mean exam scores of 3 points or between visits to progressive farmers of 6 per cent could be accidental, rather than representing true differences between the groups compared.  In turning to the computation and interpretation of this statistic we will be repeating many statistical principles introduced in connection with the chi-square test of statistical significance.

Computation of the t Statistic: The Tanzanian Election Study of 1970 included a survey of 1,544 potential voters scattered across thirteen districts of the country.  Of importance to the scholars designing this survey was the extent to which Tanzanian citizens are actively and equally involved in the political life of the nation.  One indicator of active participation is whether citizens make effective use of different sources of information about the election and their parliamentary representatives.  Responses to questions about the use of newspapers, the radio and local meetings as information souces were summarised as a media use score.  Each respondent was given a media use score, and the average media use of a particular region could then be computed.

The survey includes as well data from local leaders, including data on media use which was scored similarly to the data collected from the voters.

These two data sets permit the investigation of an important issue in a socialist
democracy, the extent to which "local TANU leaders within each community ...
monopolize the benefits of political participation for their own personal and
family advantage -- a danger supposedly increased because of the one-party
structure of the Tanzanian political system." (17, pp. 322-323)

One test of this issue is whether leaders monopolise information sources.
Thus we can compare the average use of media by the citizens with the average use
of media by the local leaders.  For a cluster of upcountry urban districts in
Tanzania it does appear that leaders make more use of information sources than
do non-leaders.  The mean score of media use by local leaders is 8.0, by the voters
it is 5.7.  This finding is particularly important because in these election districts
the education level of leaders is lower than that of followers (28% per cent of the
leaders have five or more years of education, compared to 41% per cent of the followers
with this much education).  (17, p. 324)

Perhaps the fear that local leaders are monopolising the benefits of
political participation and information is well-founded.  Such a conclusion would
be hasty, however.  Is the difference between a mean of 8.0 and a mean of 5.7
really that great?  Might it not be an accident that the leaders use the media a
bit more than the followers?

To return to our statistical language, is the difference between the
two means statistically significant?  Using very much the same logic as applied
with the $X^2$, we can answer this question.  The formula for the comparison of means
test is:

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{s.d._1^2}{N_1} + \dfrac{s.d._2^2}{N_2}}}$$

where $\overline{X}_1$ = mean of the first set of scores;

$\overline{X}_2$ = mean of second set of scores;

$s.d._1$ = standard deviation for the first set of scores; and

$s.d._2$ = standard deviation for the second set of scores.

It is convenient to designate the set of scores with the larger mean
one and the set with the smaller mean two.  Otherwise the calculation produces a

negative figure, though this presents no problem as long as one remembers that the minus sign does not matter.

The relevant data from the Tanzanian Election Study are:

Media Use During The
Tanzanian Election

|  | Local Leaders | Followers |
|---|---|---|
| Mean | 8.0 | 5.7 |
| Standard Deviation | 3.4 | 3.7 |
| Number | 29 | 167 |

When these figures are entered into the formula, it becomes:

$$t = \frac{8.0 - 5.7}{\sqrt{\dfrac{3.4^2}{29} + \dfrac{3.7^2}{167}}}$$

The calculations:

$$t = \frac{2.3}{\sqrt{\dfrac{11.56}{29} + \dfrac{13.69}{167}}} = \frac{2.3}{\sqrt{.40 + .08}} = \frac{2.3}{\sqrt{.48}} = 3.33$$

The t statistic is interpreted in a manner similar to the interpretation of the chi-square. Statisticians have worked out the distribution of t statistic values for different degrees of freedom. Computing the degrees of freedom when means are being compared is more complicated than computing it for a chi-square table, and will not be reviewed here. It often is not necessary, for when the sample is very large there is no need to compute the degrees of freedom as the distribution of t values remains nearly the same.

Usually a table used to interpret a t statistic has a row marked $\infty$, indicating the row to be used for any large sample. This is the row we can use to interpret the t statistic of 3.33.

Distribution of t statistic

|  | .10 | .05 | .01 | .005 | .0005 |
|---|---|---|---|---|---|
| $\infty$ | 1.282 | 1.645 | 2.326 | 2.576 | 3.291 |

Because 3.33 is larger than the highest value of t in this distribution, we conclude that the difference between a mean of 5.7 and a mean of 8.0 is statistically

significant at greater than the .0005 level. Stated in familiar language, in less than 5 out of 10,000 cases could we expect a mean difference as great as the one found to have happened by chance. We are relatively confident that the local leaders do use the mass media more frequently than do the followers.

Note that the t statistic, just as chi-square, tells us nothing about the magnitude of the finding. Whether one thinks that a mean difference of 2.3 is important politically is a very different judgment from whether the difference could have happened by chance. The statistic tells us the probability that the difference found is a true difference. It does not tell us what political significance to attach to the difference.

## Conclusions

Not all ideas associated with tests of statistical significance have been introduced in this chapter. For instance, statisticians and researchers often make use of one-tailed vs. two-tailed tests of significance. Moreover, there are tests of significance which can deal with more complicated data than two-variable tables or comparisons between two means. The F statistic, for example, reports the level of statistical significance associated with the differences across several means, not just the difference between two means.

The more advanced ideas associated with statistical significance can be reviewed in many good statistical textbooks. The goal in this chapter has been to review only such issues as are necessary for a basic conceptual understanding of what statistical significance means. It is not an easy concept. Yet the effective use of research writing often depends on an understanding of statistical significance. By way of summary the following points can be repeated.

1.  A statistic which measures significance or confidence levels can take many different values. A chi-square can be 1.2 or 4.9 or 8.7 or 15.1 or many other values.

2.  Statisticians know a great deal about the distribution of values which a statistic can take.

3.  They use this knowledge to estimate how hard or how easy it is to obtain a given value with a given set of data.

4.  If it is hard to obtain the value actually obtained, then they reason that the pattern in the data probably reflects true conditions.

5. If it is easy to obtain the value actually obtained, then they reason that the pattern should not be trusted; it could easily be the result of accidents or errors and not really be representative of the true conditions.

6. Statisticians translate how hard or how easy it is to obtain a given value into a probability statement. The harder it is to obtain a given value, the lower the probability that the finding in the data is an accident. The easier it is to obtain the value, the greater the probability that the finding is an accident.

7. Researchers use these probabilities to either accept their findings as probably true or to reject them as probably false.

APPENDIX

SURVEY RESEARCH AND DEVELOPMENT ADMINISTRATION:
PROBLEMS IN RESEARCH DESIGN

Preface

The material reproduced in this appendix was initially prepared as a
discussion paper for the Conference on Administrative Development in East Africa,
held in Arusha, Tanzania in September 1971.

It was written for practicing researchers, especially those engaged
in development administration research. It differs in level and in tone from
the chapters of this textbook, written primarily for beginning students in
social science methodology and for consumers of social science writing.

This essay does, however, raise a critical issue in survey research:
the correct identification of the unit of analysis in a study. And it reviews
some methodological ideas about the measurement of group properties. It is
included as an appendix for the readers who would like to go beyond the basic
materials on survey research which were presented in Chapters Two through Five.

A few decades ago, when systematic survey research was getting under-
way, social scientists were more strongly influenced by the experimental method
than they are today. Among other things this led to attempts to formulate
rigorous research designs, with the word rigorous having a very special and
narrow meaning. The survey researcher was enjoined to formulate a theory
complete with axioms and postulates. From this theory **he** would **derive** testable
hypotheses, and these hypotheses would suggest exact operations and measurements.
Each item in the survey instrument could then be traced directly to a theoretical
proposition. The analysis stage was straight-forward: apply the data to the
theory. Do the data falsify the hypotheses? If not, the theory is tentatively
confirmed; if so, the theory is rejected.

This notion of research design is not presently in favour. Because
the requirements were nearly impossible to meet, the criteria were progressively
relaxed until only the shell of the deductive approach remained. Field researchers
quickly learned that neither the research enterprise nor their own intellectual

journey could be brought under sufficient control to permit a step by step move from theory to hypotheses, to design, to test, to inference. It is no exaggeration to observe that in most survey work the final analysis reflects more the investigator's current intellectual interests than it does his intellectual interests when the study was designed, normally several years earlier. Carrying out research is itself a learning experience, and it would be an insensitive scholar who failed to learn as he went along. We well know that most survey data can serve more than one theoretical purpose, and it should not surprise us that the purpose to which the investigator puts his data need not be the purpose for which they were collected. Specific hypotheses emerge from the data as well as suggest the data to be collected.

In rejecting the criteria of the experimentalist, the survey researcher has not taken care to clarify the criteria which should replace them. Perhaps this is because the rejection has been implicit rather than overt. Whatever the reason, it is an unfortunate development. Survey researchers rush into the field with questions borrowed from other studies or with questions that sound interesting. This has two consequences, both of which should be avoided.

First, it increases the frustration ratio. I define the frustration ratio in terms of the number of times the investigator says to himself, "if only I had asked...". Poor designs lead to much frustration at the data analysis stage because (1) the wrong people were interviewed; (2) the right people were interviewed but in the wrong numbers and thus there are too few cases of what turn out to be theoretically important categories; (3) the right numbers of the right people were interviewed, but they were asked the wrong questions. Some of each mistake is unavoidable, **but** too much of any one ruins the research. There are no automatic rules to tell the scholar who to include in his sample and what to ask them, though I believe it is possible to suggest guidelines for research design which will reduce the frustration ratio without at the same time imposing the unrealistic demands of the experimentalists.

The second consequence of inadequate research criteria is increase in the cost index. This index is the cost in research resources per finding. Poor designs generate high costs and **few** findings; good designs reverse this. Costs need to be measured broadly, that is, to include much more than the money

and time of the researcher himself. Certainly they should include the burdens
placed on the respondents and on the host country, a particularly important
consideration when planning survey research on issues of development adminis-
tration. Costs should also include the spoilage effect, which follows from
saturating a particular research site or from **anta**gonising respondents. A
study which produces little in the way of significant findings but which uses
up the research potential of a particular site (such as Parliament, or a critical
training school) is a costly study indeed. And because the costs must be borne
by the entire research community, this community should cooperate in policing
its members. Standards evolve, and studies which absorb research resources
yet produce insignificant results should be terminated as early as possible.

   If replacing unrealistic criteria with no criteria is inappropriate,
what criteria should guide the construction of research designs when the major
tool is the social survey? The criteria I consider most pertinent relate to the
choice of what you intend to study. Perhaps this sounds super**fluous,** but I
think it is not. Very many studies are flawed by a failure to specify the
object of analysis. Someone decides to study administrative responsibility but
neglects to carefully consider whether he has in mind responsibility to a pro-
fessional code of ethics, to the public at large, to the elected representatives
of the public, to superiors in the administrative hierarchy or to some goal
such as equity or development. He neglects as well to specify whether adminis-
trative responsibility is best considered an attribute of the individual
administrator, an attribute of a department or bureau, an attribute of a
ministry, or an attribute of the entire civil **service;** and the choice as to
level of analysis has a great deal to do with sampling and measurement strategy,
as we see shortly.

   There are two related issues in clearly stating what one wants to
study. One issue refers to formulating concepts and the other to selecting
the proper unit of analysis. Concept formation is a sufficiently complex issue
that little can be said at this time. But it might be noted that clarifying
the concepts one wishes to study is very different from specifying hypotheses
about variables; the former can be accomplished even if the latter appears
unrealistic. Consider the concept 'career pattern': should it include
motivational traits of individual civil servants, the opportunity structure of
the civil service, pre-career experiences such as training and amount of schooling,

rules and criteria of promotions?  The dimensions included in the concept
depend of course on the research objectives, and these objectives as well as
the concepts themselves can be carefully considered well before constructing
items for a questionnaire or before beginning to interview.  It is surprising,
however, that even such simple steps as writing short working essays about
major concepts in the study seldom precede the construction of measures.

Equally critical is to take care to specify the unit of analysis.
There are three simple possibilities: **a phenomenon can be investigated on its**
own level of analysis, in terms of      **its**      unit members or in terms of
what it is a unit of.  Any two or all three of the approaches can be mixed,
but mixed strategies require great care.  Specific measures have to be matched
with each component of the overall approach.

## Levels of Analysis

Selecting the level of analysis can be simply illustrated by
considering only three possible units of analysis: individual administrators,
administrative work-groups or a ministry.  Each can be thought of as the thing
to be studied, and each can be the source of information and data, thus:

|  |  | You Intend to Study: | | |
|---|---|---|---|---|
|  |  | Individual Administrators | Administrative Work-Groups | A Ministry |
|  | Individual Administrators | a | b | b |
| You Collect Information From | Administrative Work-Groups | c | a | b |
|  | A ministry | c | c | a |

The  a cells indicate designs in which something is studied at its
own level of analysis.  The performance of extension agents is analysed with
reference to their level of training, their career motivations, their wages,
etc.-- administrators studied in terms of individual characteristics.  The
internal organisation of work-groups is analysed with reference to the informal
authority structure, the communication network, cohesion as a group --the
work-group studied in terms of group traits.  A ministry's political role is

analysed with reference to its budget, its operating policies, its legislative
mandate -- the ministry studied in terms of its own attributes.

The b cells indicate designs in which something is studied in terms
of its constituent members. A work-group and a ministry are made up of
individual administrators, and can be analysed in terms of their membership.
The cohesion and efficiency of a work-group is analysed in terms of its ethnic
composition or the prior experience of its members. Here the group is charac-
terised with respect to traits of its individual members. A ministry can be
similarly studied, as when its political role is analysed with reference to
the ideological leanings of its members or with reference to the interplay
of personal ambitions within the ministry. The simple diagram also illustrates
that organisations are composed of sub-units other than their individual members;
thus a ministry is composed of work-groups (and many other types of sub-units,
such as committees, bureaus, departments, etc.). Even the work-group has sub-
units; these may be formally appointed such as committees or more the result
of informal processes such as the formation of cliques or factions. A study
of the organisational complexity of a ministry might require information about
the number and responsibilities of semi-autonomous work-groups. Here the
ministry would be . :.. 'sed with respect to a set of sub-units which themselves
were identifiable groups in the organisation.

The  c  cells indicate designs in which something is studied in terms
of the more inclusive unit to which it belongs. We are familiar with cases in
which individuals are studied in terms of traits of the groups to which they
belong. **A member of a**work-group in which everyone has approximately the same
formal education is likely to be forthright in expressing himself; a chairman
of a work-group split by factions develops arbitration skills. An important
variation on this type of research is well-..own as the compositional effect.
The distribution of a trait in the group has an independent effect on how that
trait relates to other variables. Consider the proposition that female
committee members tend to play a more retiring role than male members, and the
finding that females are no more nor no less retiring than male members if
the committee is composed of half females and half males. Here the composition
of the group with respect to sex has an effect on the relationship between
sex and contribution to the work of the committee.

The same logic can be applied to studying other types of sub-units. The work-groups of a ministry with a coherent programme and a definite chain of command might be more effective than work-groups in another ministry which lacks much of a programme and has an unclear chain of command. Here the information about the ministry is used to analyse the behaviour of the sub-group.

The examples of the three approaches illustrated in the diagram could be multiplied many-fold, but the point needs little elaboration. An effective way to introduce rigour into research designs is to be completely explicit about the unit one wishes to make statements about and intends to collect information about, and to recognise the implications of the simple fact that these need not be the same unit.

## Measuring Group Properties

It has been said of sociology and political science that unless attention is paid to the structures of society, much quantitative analysis is nothing other than aggregate psychology. (See 3, pp. 517-527.) Traits of individuals are being added up in order to make (dubious) statements about the society, that is, societies are explained with reference to distributions of individual motives, beliefs or demographic characteristics. Examples are legion. The number of persons who score high on a modernity scale are compared to the number who score low, and a judgement is made about national development. The persons who are defferential to political authorities are counted, and a statement is made about national political culture. The number of innovative farmers are contrasted with the number who cling to traditional practices, and conclusions are drawn about patterns of innovation. The class structure of society is analysed with data reporting how respondents view subjective social status. The charge is made that much of what passes for analysis of groups, institutions, structures and patterns of relationships consists of little more than the aggregated properties of members of social units.

Part of the skepticism with which survey techniques were initially received related to fears that surveys lead to methodological individualism, to studies which have the individual and never the society as the unit of analysis. These fears were exaggerated, but not without foundation.

Probability sampling seeks out isolated individuals; the respondents are chosen
according to random procedures which make a virtue of the fact that the selection
of one respondent does not alter the probability that any other given respondent
will be chosen.  Interviews and questionnaires are administered to a single
individual, and care is taken that his responses are not biased by interactions
with friends or associates.  Yet the methodological individualism of survey
research was never as excessive as its avid critics claimed.  From the very
beginning **sub**-group comparisons and correlational techniques allowed for state-
ments about social structures, even if not for detailed analysis of relationships
among people.  However, a shift from individuals to groups as the units of
analysis, and especially to networks of relations among individuals, was delayed
in sociological analysis.  And it was a delay unnecessary given the tools of
survey research.  There is nothing inherently individualistic about survey
research.

Why do we emphasise studying administration at the group level?  Are
not programmes and policies carried out by administrators, that is, by individuals?
If so, why is it not sufficient simply to study those individuals, what they are
attempting and how well they are succeeding?  Although there is merit to these
questions, and administration is the behaviour of individuals as they carry out
their tasks, reflection will show why the administrative process can never be
conceptualised or studied with reference solely to individual administrators.

Authority structures, incentive systems and communication networks
are examples familiar to any student of administration as important unit, rather
than individual, properties.  They are also familiar as things which can be
studied with a variety of methodologies, including documentary research, obser-
vation, informal interviewing, structured questionnai**res** and analysis of
statistical data.  This latter point is important.  For though my emphasis in
this paper is with survey techniques, the more general issue of unit properties
encompasses research methodologies of all kinds.  The student of administration,
whatever  his preference for qualitative or quantitative investigations, needs
to consider the conceptual problems of explaining individual administrators in
terms of the groups (bureaus, agencies, ministries, etc.) to which they belong,
and of explaining administrative units in terms of the individuals who make
them up.  The remainder of the paper presents a more formal statement of the
different types of unit properties.  Some attention is given to how these
properties are measured, and emphasis is placed on survey techniques; but it
should not be thought that unit properties are of interest only to the survey
researcher.

Paul Lazarsfeld and his colleagues at Columbia University pioneered in applying survey techniques to the study of groups or collectivities. (For a useful summary statement, see 13, pp.499-516.) The nomenclature used here, though differing from Lazarsfeld's, owes much to his formulation. Lazarsfeld distinguished between members and collectives; collectives are such things as associations, departments, tribes, social classes and societies and the members are the individuals who compose them. But Lazarsfeld's distinction also emphasises that the members of a collective are more than individuals; cliques are members of groups, departments are members of a ministry, clans are members of a tribe, etc. We saw the importance of this earlier when attention was drawn to the level of analysis problem. We see it again in considering how to measure group properties, since several important measures require first collecting information from member units. And though the illustrations will be limited to individuals as group members, the logic could equally well be applied to other types of sub-units.

There are five types of group properties (See 7.):

| Types of Group Properties | Information From |
|---|---|
| Distributive | Members (sub-units) |
| Relational | Members (sub-units) |
| Structural | Members (sub-units) |
| Integral | Group Itself |
| Contextual | Environment of Group |

Distributive Properties: Distributive properties are obtained by performing a mathematical operation on attributes of group members. This is so common in administrative research that often what is actually a statement about the group is mistaken for propositions about individuals. An author who writes that "six out of seven civil secretary posts are held by Africans" is using a proportion to describe the extent of Africanisation, and by comparing it to a previous proportion is able to analyse the pace at which Africans replace European civil servants. When the civil service is characterised in terms of the median age of its members or the average level of education obtained, a measure of central tendency is applied. Proportions, measures of dispersion and of central tendencies, and ratios are common vehicles for characterising groups in terms of the distribution of some trait of their members.

We are familiar with measures of age, years of training, level of education, length of service, ethnicity and race, and so forth. But equally important are measures of attitudinal traits of group members. It is not uncommon to find

observations such as the following: "There was widespread resentment by civil servants against many ministers for their inadequate defense of the service" or **"this** reverence for regulations creates in some civil servants a tendency to regard themselves merely as instruments for putting the regulations into effect". Terms such as resentment or reverence identify individual attributes, but the terms widespread and some suggest that these traits are distributed in such a manner as to affect the performance of the civil service. The civil service is what is being **analysed.**

The usefulness of survey techniques is readily apparent. Surveys allow the investigator to determine relevant traits of each group member (or of a large enough sample to permit reliable estimates). From these data, proportions, ratios, measures of central tendency and of dispersion can be calculated, and thus the group itself described and compared with other groups.

Relational Properties: When we say that the members of a work-group appear to get along well or that an agency is ridden with tension, a group is being characterised in terms of the relationships which hold among the members. Relational properties cannot be investigated by concentrating on single individuals. Harmony, tension, friendship patterns, cooperation, compromise and similar group traits derive from the interactions between or among individuals. This distinguishes relational from distributive properties. In the latter, measures are reducible to traits of individuals whereas in the former they are minimally reducible to pairs of individuals.

Consider the following statement taken from a review of personnel management in Kenya:

> Excessive reliance upon formal authority of the office may lead to cold, rigid relationships which impede task accomplishment. On the other hand, the highly personal leader may become too closely involved in the work of his subordinates, thereby losing his ability for critical judgment and effective command in situations which require these. (21, p.163)

Here it is the relationships which characterise interactions between officials at different status levels which are thought to account for task accomplishment or effective command. The important fact to note is that phenomena such as "cold, rigid relationships" vs. "personal... involved" ones are properties of administrative units. That is, there are prevailing patterns of relationships in any group which have a bearing on performance.

Survey techniques are relevant to identifying relational properties,
and thus to testing hypotheses which make assumptions about the relations
which hold within or between administrative agencies. One widely used
technique is sociometric research. The members of a group are asked to name
their friends, who they go to for advice, who they most respect, how often
and in what circumstances they see fellow workers, who they share information
with and so forth. If the members of a group are asked these questions it
is comparatively easy to identify the relational properties.

Structural Properties: The analysis of structural properties is indispensable
in the study of administration. Indeed, terms such as hierarchy, status, role,
differentiation and organisation are used to define bureaucracies and bureau-
cratic processes. These terms point toward the structures of the civil service
or an agency in it. Perhaps most familiar to students of administration is
the authority structure by which is commonly meant the arrangements which determine
which role-occupants issue what sorts of commands and which role-occupants
obey what sorts of commands. When we read that authority is concentrated or
dispersed, is integrated or segmented, is secure or fragile, we recognise that
a structural property is being described.

Structures are interesting to us because they define the flow of
transactions and interactions in a bureaucracy. When, for instance, the status
structure is described the investigator knows something about patterns of
deference relationships. When the authority structure is described the
investigator knows something about patterns of directives and compliances.

The structural properties of a group can be more or less lasting.
The most permanent features of the group, especially of agencies such as
marketing boards and other administrative units, are usually identified in
its organisational chart, which is actually a representation of selected
structural properties of the group. The lines and boxes so familiar to the
reader of organisational charts simply identify the prescribed authority or
communication relationships expected to hold among the various roles of the
organisation. The chart itself of course is not a structural property of the
group, but it does permit us to identify such structures as, for instance,
the span of control.

Not all structures have been formalised and thus pictured in organisation charts. Certainly one would never find decision-structures outlined in the handbook of procedures. Yet decision-structures are very real and often very permanent properties of groups which continually must choose among policy alternatives. To know that a marketing board, for instance, has majority and minority voting blocs is to know something of more than passing interest about how it conducts its business.

There are also what we frequently term informal structures, which refer to the patterned interactions which are not sanctioned in official procedures but **which** nevertheless characterise most groups. This is what is meant by such comments as "the real way to get something accomplished around here is to work through the administrative secretary" or "important messages come directly from the field and usually by-pass the provincial headquarters". These comments identify leadership structures or communication structures which are often more critical to the functioning of the administrative apparatus than are the formalised structures of authority or communication.

Structures tell us a great deal about how people in groups relate to each other, but it is important to distinguish between relational and structural properties, though this distinction is not always easy to make. One difference is that relational properties can be reduced to dyads (that is, to pairs of individuals); structural properties cannot be similarly reduced. That is, it can be said of a group that it is generally harmonious without every member necessarily being cooperative and friendly with every other member. But it cannot be said of a group that its authority structure is pyramidical without making an assumption about the standing of every member (role-occupant) as regards the exercise of authority.

Another way in which to consider the difference between relational and structural properties is to think in terms of role-socialisation. When a new recruit joins an agency he often is instructed, formally or informally, in the duties and privileges associated with his role as well as the duties and privileges associated with other roles in the organisation. In effect he is being introduced to the structure of the group, or at least to some parts of it. Relations, in contrast, are not taught the new recruit, though due note may be made that a spirit of cooperation and harmony are expected of employees. The types of **relationships** suggested by the term relational properties

are those which emerge from the exchanges taking place between pairs of individuals. Structural properties, in contrast, suggest that certain forms of interaction are imposed on the group by virtue of the expectations and obligations associated with its roles.

Structural properties can be identified in a variety of ways, including simply looking at the organisational chart or asking a well-placed informant to describe the communication network, the status system or the author-**ity roles.** Survey techniques are also useful in studying structural properties. Thus, for instance, if no formal records are kept of voting alignments in a committee or council or board, it may be necessary to ask people with whom they regularly side if the research design calls for identifi- cation of factions. Or perhaps the investigation cannot easily proceed unless the informal structures of a group are identified. Again survey techniques are relevant, as when the members of an agency are asked to describe the flow **of information in** the group. It may not be possible to decide whether the actual communication structure is a fork (everyone sending messages to a central position which then communicates them to superiors) or a wheel (everyone communicating with everyone else), for instance, without asking persons about the number of messages they send and receive.

Sociometric analysis can be used to describe structural as well as relational properties. For instance, many of the administrative units in which we are interested are made up of equals, in that no one has formal authority over anyone else. Many committees, boards, councils and other groups have a nominal chairman who convenes meetings but who has no special authority over other members. Yet we know that groups nearly always evolve some type of leadership structure. Sociometric techniques are valuable in describing the leadership structure of a committee of equals. The researcher might ask each member to tell who in the group he most respects, who he turns to for advice, who provides the ideas which seem to lead to action and so forth. The responses to such questions would permit construction of the leadership structure. Thus, perhaps, one leader emerges for certain issues (matters of public relations) and another leader emerges for other issues (matters of internal importance to the group). In such a case, there would be a dual-leadership structure.

We simply wish to stress that survey methodologies should not be ignored even if the study advances hypotheses about the structural properties

of groups.  Indeed, being indifferent to a survey of the members may lead
to wrong conclusions if it is the case that the formal and visible structural
properties are very different from the informal and hidden ones.

Integral Properties:  Certain group properties cannot be reduced or otherwise
decomposed with reference to attributes or interactions of individual members.
These properties we call the integral properties of the group.  They belong to
the group and only to the group.  A marketing board is said to be young or old,
a reference to the length of time it has been in existence  and not to the age
of its members.  Indeed, a young board can have old members and vice versa.  A
department's budget, its growth rate or the charter which brought it into existence
are integral properties; they are independent of any attribute of individual
members.

It is a common observation about African civil services that under
colonial rule they were systems of control but that since independence they have
become systems of development.  Although authors who argue this point are not
al ʼs careful to specify what is meant, it seems fair to infer that they are
describing the basic policy orientation which differentiates the civil service
before and after independence.  Such would be an integral property, especially
if it is the result of legislation which attaches one or the other general set
of obligations to the civil service as a whole.  Other types of group properties
of course would be affected.  Thus, for instance, the proportion of law enforce-
ment officers to development technicians (distributional property) and the
appropriate conduct when dealing with constituencies (relational property)
would be affected if basic policy changes were made.  But the policy to main-
tain order or to develop would itself be an integral property of the civil
service.

Further examples of integral properties are an agency's criteria for
recruitment and promotion, its development plan, its rules for making decisions
(for instance, a committee rule to be guided by a plurality rather than a
majority), or its boundaries (as when a development cor ʼration is restricted
by law to the rural sector).  Criteria, plans, rules and boundaries are things
which characterise the group and not the individuals in the group, though of
course they influence individual behaviour and relations.

Survey techniques are less useful than other data sources if the research design requires measures of integral properties. Statistical series data might be used to describe growth rates; the public record could provide information about recruitment and promotion standards; the policy of a ministry could be observed in the written record or inferred from statutes and enactments. However, there are instances in which survey techniques might be used even with respect to integral properties.

To understand this it is necessary to distinguish between treating interviewees as respondents and as informants. Much as an anthropologist uses a well-placed informant to describe village life, a student of administrative development can use a well-placed informant to describe an agency or a bureau. That is, the interviewees would not be asked about their **age but the age of the agency, not be** asked about their training but about the training programme of the agency, not be asked their own views but about the policy of the agency, not be asked about their recruitment but about the standards of recruitment of the agency, and so forth.

There are dangers inherent in this procedure: not all informants are equally well informed (or equally open or honest). But using more than one informant and cross-checking the information can often detect biases and misinformation. Another check is to ask an informant a few questions about group properties on which reliable independent data are available, and then match his information against the independent evidence.

Dangers notwithstanding, certain integral properties can be identified only with quasi-survey methods. Many integral properties are not part of the public record or are difficult and expensive to discover. The goals of an agency set forth in enacting legislation, for instance, may be a pale (or even misleading) reflection of the operating policies. Reliable statistics on the size, wealth or growth rate of a bureau may not be available. In many instances then, an informant may provide better and cheaper data than alternative sources.

Contextual Properties: Administrative groups can be analysed in terms of the larger environment of which they are a part. When a characteristic of that environment is attributed to the group we define the characteristic as a contextual property. Such procedures are common in development administration research. The very label 'development' implies that the administrative

apparatus differs depending on whether its national environment is that of
an industrialised modern society or that of a non-industrial more traditional
society.  We also refer to contextual properties when we say that an agency
is located in the rural sector, that it carries out its task in a hostile
environment, that it is constrained by the prevailing political ideology or
that it is affected by the society's class structure.

The following proposition illustrates how a contextual property of the Kenya
civil service is used in analysis:

> Because of widespread unemployment or underemployment and an
> educational system **geared** to produce white collar workers,
> there are many more claimants for posts than posts themselves.
> Thus there is political pressure from outside the bureaucracy
> to increase the size of the administration.  And civil servants
> themselves, at least at the top of the hierarchy, are responsive
> to these demands because their own power is increased as the size
> of their ministries increases.  (2, p.61)

The rapid growth rate of the civil service (integral proper $y$)
occurs because unemployment levels and the educational curriculum (contextual
properties) lead to political pressure on the civil servants (relational
property) to which some are responsive (distributional property), at least
those at the top of the civil service hierarchy (structural property).
Analysis depending on the multiple consideration of various types of group
properties is not uncommon, as the reader himself can demonstrate by checking
any general essay on development administration.

Assigning contextual properties to an administrative unit is
comparatively straightforward.  Survey methods will seldom be necessary, though
there may be times when an informant can be used.  The most frequent use of
surveys in connection with contextual properties derives from the fact that
the context of a lower level unit is often a distributive, relational,
structural or integral property of that context.  The extension service is
a sub-unit of the ministry, the ministry can be viewed as context within which
the extension service carries out its duties.  The expertise of the ministry
(a distributional property), the patterns of cooperation throughout the
ministry (relational proper $y$), the lines of authority (structural property)
and the overall size characterise the ministry along relevant dimensions
which then are assigned as contextual properties to the extension service.

Implications and Conclusions

The most obvious benefits of distinguishing among these different group properties have to do with research design. In the first place, it reminds us that many of the important units of action in the administrative system are groups. Indeed, this is true of the entire political system. Courts, review panels, legislative and executive committees, city councils, juries, working parties, voluntary associations, marketing boards and party cells are just a few among the many types of groups which have or affect powers of government. Common language is full of references to political and administrative groups, to how they are doing their jobs and to who is benefiting from their decisions. The sooner students of politics and administration devise methods of analysing groups, the sooner our social theories will be adequate to explaining what goes on in political life.

Moreover, even if the investigator is more interested in individuals than groups, in the administrator or Member of Parliament or union leader or party member, much of what must be understood about the behaviour of individuals in politics can be explored with information about the groups to which they belong. To use the language suggested earlier, to explain something in terms of what it is a unit of nearly always requires some sophistication in measuring group properties.

Designing research benefits in a second way from the five-part classification of group properties. Decisions about the type of data to collect, about sampling procedures and about the form of data collection necessarily differ depending on the type of properties relevant to the research objectives. Only a few ramifications of this can be explored here, but if the frustration ratio and the cost index are to be reduced, it will require research designs which specify the objects of study and which clarify their relevant properties.

If the researcher expects relational properties such as cohesion or tension to be relevant to administrative performance then it is necessary to design a study which permits him to locate individuals in terms of the relations they have with other group members. A failure at the design stage will result in no measure of cohesion or tension at the analysis stage. Similarly, if he expects that communication patterns or authority structures are critical variables, then it is necessary to design a study which will produce measures that differentiate agencies in terms of such structural variables.

The following paragraph is taken from an essay by R.J. Ouko,
Minister of Finance and Administration for the East African Community; it
outlines an important proposition relevant to development administration.

> The spirit of the civil service is also affected by the presence
> of expatriate officers. Many of these have set an example of hard
> work and devotion to duty and identify strongly with the public
> service, but the tendency of some to operate outside the formal
> chain of command can create difficulties. African officers become
> frustrated when they see the ease with which some expatriate
> officers get access to top officials in contravention of the
> established code. This type of administrative behaviour has a
> high cost in lower morale among African officers. The advantages
> gained through such informal administrative relationships must
> always be weighed aga٦st their costs in c eating administrative
> tensions between expatriates and their African co-equals, thus
> impairing the g٦ eral growth of civil service loyalty. (21,p.163)

The elaborate proposition outlined in  his paragraph depends on
assumptions about various types of group properties, and the proposition
could not be tested unless the research design permitted measures of these
properties. The actual wording of items in an interview or questionnaire
depends on a clear view of what type of property is being measured. The
formally established channels of communication are a structural property,
and from a study of these one would learn which levels of civil servants lack
the formal right of direct access to, say, the Permanent Secretary. To
ascertain this structural property, one might treat the higher civil servants
as informants and ask them, "Which positions in this ministry carry with them
the right of direct access to the Permanent Secretary?" Then taking those
levels without such access, one would want to know whether expatriates actually
do have direct contact with the Permanent Secretary and   whether this access
is greater than that of African officers operating at the same level.

Access is a relational property. To study it we would ask each
civil servant, expatriate and African, "Are you ever able to communicate about
your work directly and personally with the Permanent Secretary or do you find
it necessary always to go through someone else?" (If the former) "How often
would you say that you are able to communicate directly and personally with
the Permanent Secretary?" As a check on this information, the Permanent
Secretary might be asked, "Which members if this ministry often communicate
with you personally and directly about their work?" By using the structural
property of formal channels as an indicator of prescribed distance from the
Permanent Secretary, we next would have established whether expatriate and
African officers do in fact tend to be different with respect to the relational
property of access.

From this point, we would then need to establish two distributional
properties and one more relational one. The African officers would be asked,
"Do you feel that expatriates tend to have more direct access to the Permanent
Secretary than African officers of the same formal rank?" and a question
probing loyalty such as, "If you were offered a slightly better paying job
in the private sector, do you think you would take it?" These questions would
establish the distribution of a perception and an attitude among African
officers. The other relational property concerns the hypothesis of tension
between African and expatriate officers and could be investigated by asking,
"Are there any particular officers in this ministry with whom you find it
difficult or unpleasant to work?" Here one would be looking for a disproportionate
number of expatriate nominations on the part of Africans and vice versa. Thus
we see that an identification of the type of group property being investigated
helps us determine the individuals to whom a question is to be put, whether
the interviewee is to be treated as an informant (when he can be mistaken) or
as a respondent (when he cannot) and whether interviews are to be treated as
single units (distributional properties) or as pairs (relational properties).

The utility of the five-part classification of unit properties is
also seen when we consider sampling strategy. This is not the place for a
complex review of sampling in survey research, but the previous discussion
indicates that differing research questions call for different samples. At
the outset the investigator must decide whether **he** wishes to sample adminis-
trators or administrative units, or to draw a multistage sample in which first
units are selected and then some proportion of members within each unit.
Analysis solely concerned with contextual or integral properties might only
sample administrative agencies. Analysis considering distributive, relational
and structural properties would have to sample **ind**ividuals within identifiable
groups. Whether the total membership would be interviewed, or only a part of
the membership, would depend on the size of the group as well as the type of
property to be identified. (Identification of most structural and relational
properties in small groups requires interviews with all members of the sampled
groups.) Pairs of individuals could also be sampling units, as in a study
investigating relation**ships** between expatriate technicians and their African
counterparts.

Of course sampling decisions are made about data other than those
collected from individuals. A study of communication structures might sample
written messages or telephone calls which pass between local, district and

- 188 -

national offices.  A study of administrative policy might sample directives
or enactments.  A study of administrative effectiveness might sample crop
production or number of licenses awarded.  But whether data are to be collected
from individuals or from some form of document, it is still necessary to know
whether the sampling unit is the agency itself.

_____

A previous paper, written with David K. Leonard, "Methodological
Notes on  uantification, Productivity and Groups in Administrative Research"
(Discussion Paper No. 110, Institute for Development Studies, University of
Nairobi, 1971), employs the five-part classification of group measures used
in the latter part of this paper.  Where relevant I have reproduced, without
citation, sections of that paper.  I am grateful to Mr. Leonard for permission
to use this material.

BIBLIOGRAPHY

1. Barkan, Joel. Personal communication. August 5, 1974.

2. Bienen, Henry. "The Economic Environment." In Goren Hyden, Robert Jackson and John J. Okumu, editors. Development Administration: The Kenyan Experience. Nairobi, Oxford University Press, 1970.

3. Coleman, James S. "Relational Analysis: The Study of Social Organizations with Survey Methods." In A. Etzioni, editor. A Sociological Reader on Complex Organizations. New York, Holt, Rinehart and Winston, 1969.

4. Court, David and Ghai, Dharam P., editors. Education, Society and Development: New Perspectives from Kenya. Nairobi, Oxford University Press, 1974.

5. Court, David and Prewitt, Kenneth. "Nation Versus Region in Kenya: a Note on Political Learning." British Journal of Political Science. 4, pp. 109-120.

6. Currie, Janice. Social Characteristics of Secondary School Students and Their Distribution in the Occupational Structure of Uganda. Paper No. 17. East African Universities Social Science Council, Eighth Annual Conference, held in Nairobi, Kenya, on December 19-23, 1972.

7. Eulau, Heinz and Prewitt, Kenneth. Labyrinths of Democracy. Indianapolis, Bobbs-Merrill Co., 1973.

8. Gachuhi, J. Mugo, Chege, Fred E. and Ascroft, Joseph. Kirathimo Model Village of the Kenya Red Cross Society: An Evaluation Report. Miscellaneous Paper No. 48. Institute for Development Studies, University of Nairobi, 1972.

9. International Labour Office. Employment, Incomes and Equality: A Strategy for Increasing Productive Employment in Kenya. Geneva, 1972.

10. Institute of Development Research. Ghimbi Socio-Economic Baseline Survey: Preliminary Report of Research Proceedings. Addis Ababa, Haile Selassie I University, 1974.

11. Karuga, J. Land Transactions in Kiambu. Working Paper No. 58. Institute for Development Studies, University of Nairobi, 1972.

12. Kornhauser, A. and Steatsley, P. "Questionnaire Construction and Interview Procedure." Appendix C in Claire Selltiz, et al. Research Methods in Social Relations. New York, Holt, Rinehart and Winston, 1967.

13. Lazarsfeld, Paul F. and Menzel, Herbert. "On the Relation Between Individual and Collective Properties." In A. Etzioni, editor. A Sociological Reader on Complex Organizations. New York, Holt, Rinehart and Winston, 1969.

14. Leonard, David K. and Prewitt, K. Methodological Notes on Quantification, Productivity and Groups in Administrative Research. Discussion Paper No. 110. Institute for Development Studies, University of Nairobi, 1971.

15. Mbilinyi, Marjorie J. _The Education of Girls in Tanzania_. Dar es Salaam, Institute of Education, University College, 1969.

16. Mbithi, Philip M. "Agricultural Extension as an Intervention Strategy: An Analysis of Extension Approaches." In David K. Leonard, editor. _Rural Administration in Kenya_. Nairobi, East African Literature Bureau, 1973.

17. Morris, Jon R. "The Voters' View of Elections." In Election Study Committee, University of Dar es Salaam. _Socialism and Participation: Tanzania's 1970 National Elections_. Dar es Salaam, Tanzania Publishing House, 1974.

18. Mwansasu, B.U. "The Selection of Candidates." In Election Study Committee, University of Dar es Salaam. _Socialism and Participation: Tanzania's 1970 National Elections_. Dar es Salaam, Tanzania Publishing House, 1974.

19. Nyangira, Nicholas. Relative Modernization and Public Resource Allocation in Kenya. Paper No. 56. East African Universities Social Science Council, Eighth Annual Conference, held in Nairobi, Kenya, on December 19-23, 1972.

20. O'Barr, William, Spain, David H. and Tessler, Mark A. _Survey Research in Africa: Its Applications and Limits_. Evanson, Northwestern University Press, 1973.

21. Ouko, Robert J. "Personnel Management." In Goren Hyden, Robert Jackson and John J. Okumu, editors. _Development Administration: The Kenyan Experience_. Nairobi, Oxford University Press, 1970.

22. Payne, S.L. _The Art of Asking Questions_. Princeton, Princeton University Press, 1951.

23. Wilson, Rodney J.A. The Economic Implications of Land Registration in Kenya's Smallholder Areas. Staff Paper No. 91. Institute for Development Studies, University of Nairobi, 1971.