

The Collection, Analysis, and Use of Monitoring and Evaluation Data

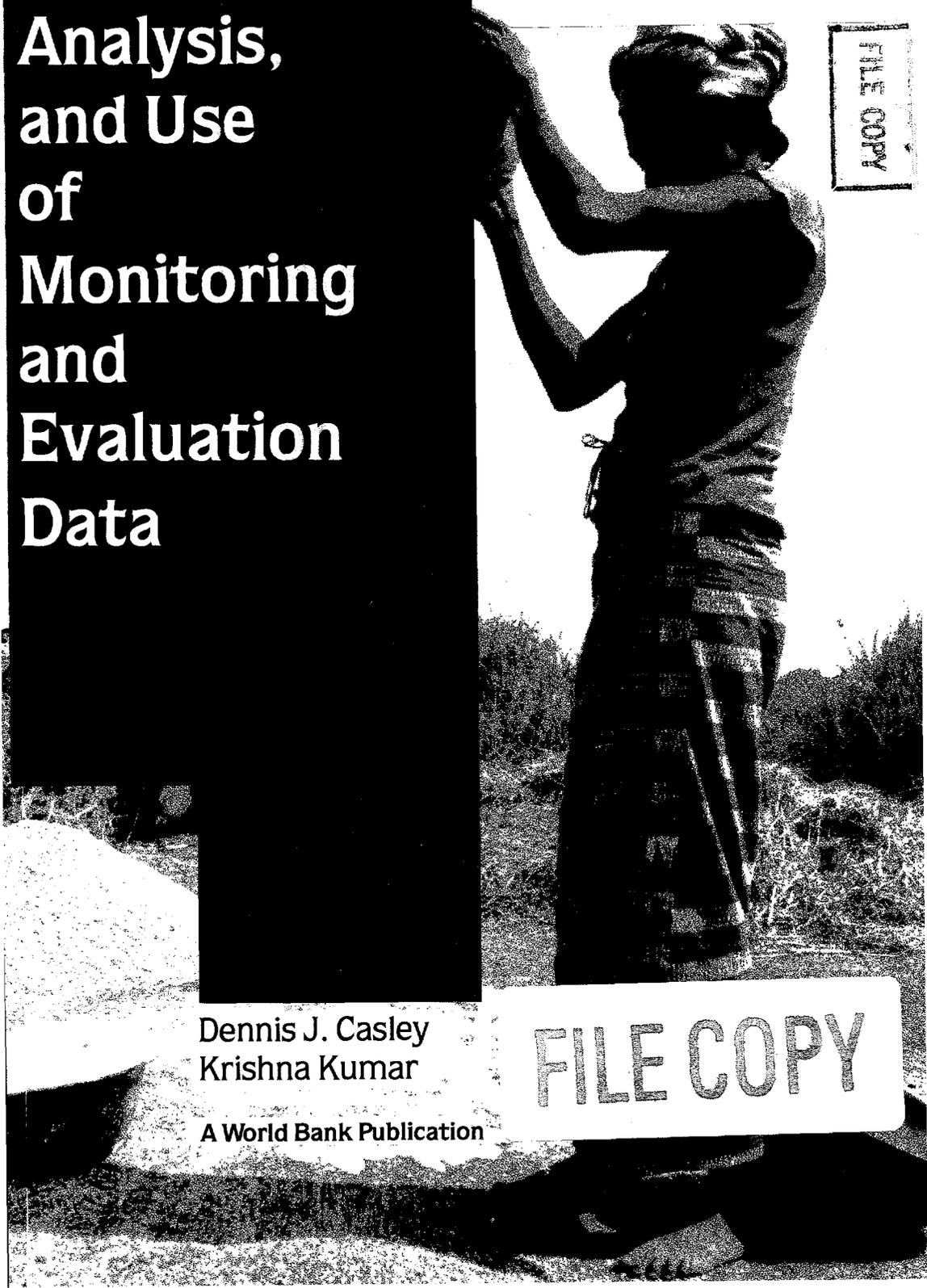
9456

FILE COPY

Dennis J. Casley
Krishna Kumar

A World Bank Publication

FILE COPY



*The Collection,
Analysis, and Use
of Monitoring
and Evaluation Data*

A Joint Study



The World Bank



International Fund
for Agricultural
Development



Food and Agriculture
Organization of the
United Nations

This book is the second in a series on the monitoring and evaluation of agriculture projects. The series includes a companion volume, *Project Monitoring and Evaluation in Agriculture*, as well as brief technical notes to be published in the World Bank Technical Papers series.

*The Collection,
Analysis, and Use
of Monitoring
and Evaluation Data*

Dennis J. Casley and Krishna Kumar

Ram C. Malhotra, editor of the series

*Published for The World Bank
The Johns Hopkins University Press
Baltimore and London*

© 1988 The International Bank for Reconstruction
and Development / THE WORLD BANK
1818 H Street, N.W., Washington, D.C. 20433, U.S.A.
All rights reserved
Manufactured in the United States of America

The Johns Hopkins University Press
Baltimore, Maryland 21211, U.S.A.

The findings, interpretations, and conclusions
expressed in this study are the results of research
supported by the World Bank, but they are entirely
those of the authors and should not be attributed
in any manner to the World Bank, to its affiliated
organizations, to members of its Board of
Executive Directors or the countries they represent;
to the International Fund for Agricultural
Development; or to the Food and Agriculture Organization
of the United Nations.

First printing March 1988
Second printing January 1989

Library of Congress Cataloging-in-Publication Data

Casley, D. J.

The collection, analysis, and use of monitoring and
evaluation data.

Bibliography: p.

Includes index.

1. Statistics. 2. Economic surveys. 3. Economic
development projects—Evaluation—Statistical methods.
4. Sampling (Statistics) I. Kumar, Krishna. II. Title.
HA29.C36 1988 001.4'2 87-46375
ISBN 0-8018-3670-7
ISBN 0-8018-3669-7 (pbk.)

Cover photograph: Winnowing grain in Chad. By A. Girod.
Courtesy Food and Agriculture Organization
of the United Nations.

Contents

<i>Acknowledgments</i>	<i>ix</i>
<i>1 Introduction</i>	<i>1</i>
Purposes of Data Gathering	2
Qualitative and Quantitative Data	3
Constraints on Options for the Collection of Data	6
Data Analysis and Interpretation	7
<i>2 Qualitative Interviewing of Individual Informants</i>	<i>10</i>
Types of Qualitative Interviews	10
Guidelines for Qualitative Interviews	15
Reliability of the Interview	21
General Respondents and Key Informants	24
Limitations of Qualitative Interviews	24
<i>3 Conducting Group Interviews</i>	<i>26</i>
Community Interviews	27
Focused Group Interviews	34
Limitations	40
<i>4 Participant Observation</i>	<i>41</i>
Conceptual Framework and Data Requirements	42
Site Selection and Timing	45
Understanding Fieldwork	46
Study Instruments	49
Minimizing Biases	51
Limitations	52
<i>5 Structured Surveys</i>	<i>54</i>
Socioeconomic Surveys: Baseline and Follow-up	54
Planning a Structured Survey	56
Data Requirements	57

Survey Design	57	
Concepts and Definitions	59	
Questionnaire Construction	64	
Interviewing the Respondents	74	
6 Sampling for Monitoring and Evaluation		76
Probability versus Informal Sampling	78	
Sample Size	83	
Single-Stage Sampling Techniques	85	
Two-Stage Sampling Techniques	92	
Sampling for Rare Events	94	
7 Measurement of Crop Production and Yields		96
Area, Production, and Yield	97	
Area Measurement	99	
Yield and Production Measurement	102	
8 Exploratory Analysis		114
Simple Graphic Examination	115	
Ordering the Data and Measures of Location	117	
Trimming and Transforming Data	123	
Measures of Dispersion	126	
Fitting Straight Lines	127	
Analysis of Residuals and First Differences	127	
Moving Averages	129	
Proportions	131	
Analysis of Subsamples	131	
9 Statistical Analysis of Data		132
Comparing Two Sample Means	133	
2 × 2 Tables	134	
Cross-Tabulations	137	
Comparing Differences from Multiple Groups: The Analysis of Variance	138	
Regression and Correlation	142	
Cautionary Comments about Correlation and Regression	145	
10 Presenting Data to the User		148
Frequency Distributions	149	
Truncation of Digits and Summary Distributions	150	
Presentation of Cross-Tabulations	155	
Presenting Percentages	159	
Graphs and Figures	161	

Attribution of Accuracy and Significance	161
Writing the Main Report	167
<i>Suggested Readings</i>	169
<i>Index</i>	171
<i>Examples</i>	
1 A Guide for a Topic-Focused Interview	12
2 An Interview Involving Sensitive Questions	16
3 Involving Participants in Community Interviews	31
4 Group Interviews for Generating Quantitative Data in Costa Rica	32
5 Introductory Remarks for a Focused Group Interview	36
6 Developing a Conceptual Framework for a Participant- Observation Study	43
7 Steps in Conducting Participant-Observer Evaluation	48
8 Observation Record Form for General Meetings of a Multipurpose Cooperative Society	49
9 Document Summary Form	50
10 A Two-Way Table for Livestock Numbers	65
11 Agricultural Research and Extension Project	66
12 A Verbatim Survey	70
13 Recent Study of Maize Yields in Zimbabwe	105
14 Bangladesh Postharvest Project	106

Acknowledgments

As was its companion, this book is a joint effort of the World Bank, the Food and Agriculture Organization of the United Nations (FAO), and the International Fund for Agricultural Development (IFAD). We particularly acknowledge the help of Ram Malhotra, the series editor, and Monica Fong of IFAD; and Chandra Arulpragasm, Clifford Morojele, and Françoise Petry of the FAO.

Chapter 6 is based on *Sampling for Monitoring and Evaluation*, a monograph by Chris Scott previously published (Washington, D.C.: World Bank, 1985); revisions were undertaken by him. Similarly, chapter 7 draws on another monograph, *Estimating Crop Production in Development Projects*, by C. Derek Poate and Dennis J. Casley (Washington, D.C.: World Bank, 1985). Vijay Verma also contributed to the revision of this chapter and provided comments on others. We thank these colleagues for their assistance.

We also wish to thank our colleagues Ronald Ng, Vinh Le-Si, and Josette Murphy for their substantial comments and assistance during the writing of the book. Diana Crowley once again provided us with invaluable assistance in researching material and producing the data for the examples. We are grateful also to many who reviewed sections of the manuscript. Finally, we thank Joy Vendryes, Workie Ketema, Susan Smith, and Michael Alloy for patiently preparing the various drafts of the manuscript.

1 | Introduction

THIS VOLUME FOCUSES on the topics in data collection, analysis, and use that were raised in its companion book, *Project Monitoring and Evaluation in Agriculture* (Baltimore, Md.: Johns Hopkins University Press, 1987). The introductory chapter summarizes these topics in the context set by the companion volume, which documented the importance of monitoring and recommended that ambitious evaluations be done only selectively.

Later chapters in this book further explain the data collection and analysis techniques referred to in the companion volume. The book is thus selective; it does not provide comprehensive coverage of all methods. Furthermore, for those methods it does cover it advocates simplicity and economy, which we believe are necessary given the limited resources of many development projects.

We emphasize qualitative interviewing methods because most monitoring and evaluation efforts will need to use them for limited, nonrandom coverage of respondents. Our examples of sample theory and sample selection, conversely, are mainly in terms of rates of adoption and similar indicators because most projects require moderate-size random samples to monitor such rates. In a similar vein, we advocate greater reliance on farmer estimates when discussing crop production and yield, for we are concerned with this issue in the context of project beneficiary responses rather than of aggregate national or regional estimates.

Finally, we make no apology for the very simple treatment of data analysis issues. We know that many of our readers, isolated at a project site, seek a source of simple calculation procedures and simple guidelines for constructing tables. And some of the simple rules frequently are broken by even sophisticated analysts with libraries of textbooks.

Because each chapter deals with a specific area of data collection, analysis, or use, it is somewhat discrete and autonomous. Although there is some logic to the chapter sequence, the book is meant to be consulted as a particular issue arises rather than read straight through. In order, the subjects covered are qualitative data collection methods; structured surveys

and sampling and crop measurement problems; preliminary, exploratory data analysis; formal analysis; and data presentation.

Purposes of Data Gathering

The data gathered for monitoring and evaluation have three purposes: description, explanation, and prediction. These purposes are not mutually exclusive; rather, there is a logical progression from the first to the last. The description of a phenomenon or process is the first step toward explaining its nature, underlying causes, relationships, and context. And prediction usually, though not always, requires both description and explanation.

Descriptive data answer questions of who, when, and where but not of how or why, which would require probing into causes. The data generated for physical and financial monitoring are the best examples of descriptive data. Although they record progress, describe the relationship between expenditure and achievement of the physical targets, and identify possible deviations from the planned course, they do not shed light on the reasons for progress or the lack of it. Most of the data from beneficiary contact monitoring are also descriptive.

An explanation requires an extrapolation of a cause and effect relationship; the investigator tries to understand why a phenomenon, process, or event occurred or did not occur. In diagnostic studies, we search for an explanation. For example, when project staff are trying to find out why farmers in the project area are not responding positively to a technical package which has proved highly successful elsewhere, they are searching for factors and conditions that shed light on this unexpected phenomenon. They are seeking to answer the question of why.

The presence of a large number of factors and conditions interacting with one another in a changing environment makes it difficult to establish a causal relationship in the study of economic and social phenomena. What we generally get is a reasonable indication of a strong association between a set of variables in a temporal sequence, which is logically justifiable. This, we suggest, is more than sufficient for most project evaluation situations.

In the past, the ambition to establish causal relationships on the basis of hard, empirical data often led to unrealistic evaluation designs which were beyond the capability of monitoring and evaluation resources and proved unsuccessful. We stress throughout the companion volume the need to have realistic expectations for evaluations.

Predictions are based on an understanding of the causes of events. Project staff are usually not involved in long-range predictions about the outcomes of a project or its components. In most cases, they do not go be-

yond making short-term projections that take into consideration a few variables.

The above primary uses of information—description, explanation, and prediction—affect the overall design, scope, and modes of data collection and the analysis. More specifically, the data gathered in monitoring and evaluation are used:

- To monitor physical and financial progress. The data required for physical and financial monitoring are usually available in project records and documents. The primary focus in this particular case is with the collation, rather than with the collection, of data, and with its analysis.
- To examine the responses of the project beneficiaries to the services and inputs being provided by the project. To do this requires structured surveys designed to answer questions on beneficiaries' knowledge of, reaction to, and future use of project inputs and services. Here the data collection methodology is the principal issue; analysis of such data as adoption rates is relatively simple.
- To study specific implementation problems facing a project with a view to diagnosing their cause and suggesting practical solutions. Such investigations, called diagnostic studies, are to be undertaken from scratch and completed within a matter of weeks rather than months so that timely, relevant information and recommendations are available. A blend of data collection methods may be appropriate, but qualitative information will almost certainly be required.
- To assess (or predict) the effects of the project on production. Two questions are raised in this regard: Was any change in levels of production or yield detected? Can such a change be attributed to project intervention?
- To assess the socioeconomic impact of the project, particularly by collecting and interpreting data on income, living standards, people's participation, and the environment. Such data raise the most complex issues of all the data requirements, which is why we advocate selectivity in conducting such studies and practical, limited objectives for them.

Qualitative and Quantitative Data

There are two types of data collection methods: qualitative and quantitative. The most obvious distinction between the two is that quantitative methods produce numerical data and qualitative methods result in information which can best be described in words. Examples of the latter are "descriptions of situations, events, people, interactions and observed be-

haviors; direct quotations from people . . . and excerpts or entire passages from documents, correspondence, records and case studies.”¹

Moreover, qualitative methods focus on the signs and symbols that decode the reality seen by the target population. This is reflected during data collection. Suppose in response to the same question two farmers give the identical answer: “If I had money, I would purchase a small tractor to plow my fields.” The interviewer in a structured survey will code the farmers’ responses identically, for both said the same thing. If qualitative methods are used, however, the investigator will be more cautious and will also examine the context and manner in which they spoke. In other words, an attempt will be made to see if the two farmers meant the same thing. Both verbal and nonverbal behavior are examined in qualitative studies in order to understand the views, attitudes, and perspectives of the respondents.

Finally, qualitative methods are iterative. There is an ongoing opportunity to revise interview protocols, guides, and observation record forms as the study progresses and new facts are brought to light. Even the underlying framework can undergo significant changes at the stage of data collection. Such iteration is not common in the collection of quantitative data. The enumerators in a structured survey, for example, are not usually expected to ask additional questions, even if they feel such questions will provide useful information. Nor are they allowed to discard an interview, even if they have a hunch that the respondent did not give candid answers.

The most widely used method for collecting quantitative data is the structured survey, which entails administering a written questionnaire to a sample of respondents. Such surveys can be done at one point in time or at various intervals. The latter are useful for discerning changes and trends in yields, productivity, and the standard of living of the target population. The advantages of structured surveys are that the interview mode and construction of questions can be standardized on the basis of experience so that the size of biases introduced by either the enumerator’s style or the respondent’s misunderstanding is controlled. Furthermore, if sampling theory is used to select the respondents, the sample results can be used to derive estimates for the whole population within known margins of probable error.

Quantitative data may be obtainable from the records of project agencies and other institutions. Statistical offices may have extensive data on file which can be recoded, aggregated, disaggregated, and reanalyzed for diagnostic studies and impact evaluations. Such data must be used with caution because serious errors can result from differences in the defini-

1. Michael Q. Patton, *Qualitative Evaluation Methods* (Beverly Hills, Calif.: Sage, 1980).

tions of main variables or in the circumstances in which the data were collected.

One method of obtaining qualitative data is to conduct in-depth interviews with individual respondents. In such interviews, the interviewer gently probes the respondent, which permits them to have a conversation in which ideas flow freely. The interviewer must, however, take elaborate notes. Such interviews can be conducted with a few well-informed persons called "key informants" or with ordinary members of the target population. Village leaders, tribal chiefs, extension workers, teachers, and local government officials generally make good key informants.

A second qualitative method is group interviews. These are usually of two types: community meetings and focused group discussions. The former are open to all adults in the community or village. They are usually well attended if sufficient notice is given. Such meetings are best conducted by a team of two or three interviewers, who address queries to the participants. Focused group discussions are sessions for a small number of invited participants, who discuss a topic among themselves. The interviewer simply stimulates the discussion and keeps it focused on a desired topic.

A third qualitative method, participant observation, involves direct, extensive observation of an activity, behavior, or relationship. Participant observation can also include qualitative interviews with the informants. The merit of this approach is that the investigator gets an inside picture of the situation as seen by the people involved.

Most information systems within projects require the collection of both quantitative and qualitative data. Choosing the right method for each particular data need is a principal responsibility of the monitoring and evaluation staff; both kinds of data have their strengths and weaknesses.

Quantitative data are obviously needed when a number, rate, or proportion related to the target population must be estimated or a variable such as crop production must be measured. Qualitative data are needed when the attitudes, beliefs, and perceptions of the target population must be known in order to understand its reactions to project services.

The decision whether to gather quantitative or qualitative data in a particular case also influences decisions regarding the scale of the survey and the procedures for selecting respondents. Qualitative methods are best used with small numbers of individuals or groups—which may well be sufficient for understanding the human perceptions and behaviors which are the main justification for a qualitative approach. Moreover, the selection of the individuals or groups may be done deliberately according to required characteristics rather than at random from the population.

If quantitative inferences about the population or a section of it must be made, however, a random selection of a somewhat larger number of

respondents dispersed throughout the population will be necessary. Many options exist as to how to design such a sample (these are introduced briefly in chapter 6). Complex sample designs may occasionally be necessary, but the design usually should be simple.

One other decision may be linked to the choice of data collection method: how broad or narrow the study should be. One may restrict the study to the collection and analysis of only the essential variables of interest. Or one may wish to interview respondents about a broad range of issues so as to be able to explore relationships and reactions that may unexpectedly turn out to be relevant.

The companion volume states that data collection for monitoring and evaluation should be limited in scope and sharply focused. The reasons for this—constraints on skills, time, and budgets—are described in the next section. Certainly the number of variables in quantitative surveys should be kept to a minimum. Data collection to supply a management information system is not meant to be done with general-purpose, multisubject surveys.

In the case of qualitative studies, there is more justification for a broad, in-depth pursuit of a subject. Even in such cases, however, the study should limit its range to the subject of concern rather than try to cover a wide range of subjects on the grounds that multiple purposes may be served within a single inquiry. Multisubject qualitative studies tend to be vaguely conducted and result in poor data.

Constraints on Options for the Collection of Data

A common constraint on the options at each stage of the process of designing a system to collect data is the level of available resources. The old saying that one should cut one's suit to fit one's cloth is apt. It is fruitless to design a data collection operation to reach a widely dispersed sample of respondents with the use of questionnaires and in-depth interviewing techniques if neither the logistic resources to manage such an operation nor the skills to collect and observe accurate data are available.

Clearly, the options and tradeoffs in the design of an operation to collect data are not reviewed and decided upon only according to criteria of theory and precision. One must have the resources to collect the data on the scale envisaged and in the time required and the resources to turn the data into useful information. Available funding is clearly important, but money alone is not the main issue. Survey management skills will be the vital—and in many cases limiting—factor, for these will be in shortest supply. If a small unit within a project has limited survey experience and skills, it should concentrate on simple monitoring, contact-type surveys rather than ambitiously embark upon a complex longitudinal study of socioeconomic variables that will be doomed to failure by poor data resulting from indifferent questionnaire design and inadequate interview-

ing skills. In such a case, this limit must be clearly recognized and understood by data users and collectors alike at the planning stage. This is the principal theme of the companion volume.

Another basic constraint on available options is the time within which the data must be collected and analyzed. This very real issue is ignored extraordinarily often. The companion volume comments on a number of surveys that collected data over several seasons for an ambitious analysis requiring a time series that outstripped the life of the project. There may be a case for such a patient endeavor if there will be users outside the management of the project. But surveys usually are justified as a means of producing information to improve the implementation of the project. We have said elsewhere that information required to help managers make decisions becomes valueless—however accurate—if it is provided after the decisions have been made. This can frequently impose a very tight time limit on a monitoring officer attempting to carry out a diagnostic study. Methodology may have to be ad hoc: there may be no time for adequate sample frame preparation; the number of respondents may have to be limited; and the interviews may have to be conducted by the monitoring officer and a couple of trusted aides rather than a bevy of field staff. Monitoring officers must be flexible and pragmatic, but this is acceptable as long as all parties are aware of the constraints that have been applied and the inferential limitations thus imposed.

The converse of course is also true. If time is not a constraint, then the senior staff should give the preparatory phases of data collection the attention and care they merit. The design of questionnaires is a prime example: insufficient time for design and a failure to test the draft are commonplace, even when time is available for these vital steps. Should a survey require the use of junior enumerators, training for them is another preparatory aspect that must not be skimmed on. However skillful the design of the questionnaire and however enthusiastic the enumerators, if they go into the field less than totally confident and with less than a total grasp of the entire range of the interview and methodology concerned, the survey will fail. Bringing the enumerators to this level takes considerably longer than many people suppose. The length of training obviously depends on the complexity of the survey. It does not take very long to train an enumerator to conduct a simple interview on the adoption of a particular recommendation, but it may take two weeks to train a limited number of enumerators to measure land areas and crop yields and to interview farmers regarding their marketing behavior.

Data Analysis and Interpretation

The companion volume emphasized repeatedly that the only justification for collecting data is that they will be used for a specific purpose that has been clearly identified and agreed upon in advance. Even if the ap-

appropriate data are collected and processed accurately, analysis and interpretation are essential if they are to be used to facilitate decisionmaking. This may seem a tautological statement, yet there are many examples of valuable data rendered valueless by an inability to turn them into usable information.

At one extreme, there is a total failure to carry out analysis if a computer is used to prepare basic, voluminous tabulations and these are passed on to managers as if they constituted information rather than merely regurgitated data from a machine. At the other extreme, there is an excess of analysis if complex statistical techniques are applied in order to fit relationship functions and distill the data down to regression coefficients and significance-level statements. Often these techniques go beyond what is required, are misapplied to data that do not meet the statistical conditions, and result in coefficients that the user does not understand.

Monitoring staff should pay much greater attention to simple exploratory analysis and to carefully presented tabular material. There are occasions when more advanced analytical techniques are demanded, but even then they should be used only if their theoretical basis and the conditions that validate their use are fully understood. We discuss these matters in chapter 9.

The decision whether computers are needed to help maintain and use the management information system is one that most monitoring and evaluation staff and their project managers have to make in these days of easily accessible, inexpensive microcomputers. Computerization should be considered carefully during project preparation as part of designing the information system; it should not be undertaken automatically—as is now frequently done. As the staffs of many current projects—particularly those involving extensive data collection—have found, computerization is almost assumed. But computerization alone does not guarantee the smooth function of data operations; it often causes delays and problems. Computerization also absorbs both scarce capital and manpower in many projects that could be devoted to other important tasks.

Recommendations on computerization can be summarized by the following statements:

- If an existing manual system works efficiently and no extensive data gathering is required, computerization is not an immediate concern.
- If report preparation is a big problem, and many revisions are needed before final approval, some word processing capability should be considered (plans for expansion to include data processing should be taken into account at this stage).
- If data will be collected extensively for several years and analyzed intensively and quantitatively, plans should be made for computeriza-

tion. In the case of survey data processing, if sample sizes are in the hundreds, a large number of variables are required, and there are many repetitive calculations, computerization provides a definite advantage.

- If computerization is contemplated, detailed plans for organizing and operating the system should be made while the information system is being designed.
- There should be provision for backup to the system in case of major breakdowns; maintenance contracts should be established with vendors and should include provisions for equipment loans.
- To meet the needs for skilled staff to operate a computerized system, the project should have a minimum of two specialists able to maintain service continuity.
- Training should be provided not only to computer staff but to other staff who use the system regularly, and this training should be a continuing process to keep trained persons up to date.

Whereas it is possible to introduce the reader to the analytical tools used to handle numerical data, the analysis of qualitative information is a different matter. Certain indicators may be derived from data that can be treated in a quasi-numerical manner. As will be seen in the following three chapters, however, most learning from qualitative interviews is obtained by writing descriptive summaries and collating and sorting these summaries into categories of response. The best advice we can offer the inexperienced in this context is the same that might be offered to a budding writer: read many examples by respectable practitioners of the art.²

2. A good starting point is Matthew B. Miles and A. Michael Huberman, *Qualitative Data Analysis: A Sourcebook of New Methods* (Beverly Hills, Calif.: Sage, 1984).

2

Qualitative Interviewing of Individual Informants

ONE OF THE MOST IMPORTANT SOURCES of information for monitoring and evaluating agriculture and rural development projects is qualitative interviews. Qualitative interviews with project participants and other key informants help in understanding the complex ecological, sociological, cultural, and other situations with which the project must deal. They can also provide an in-depth understanding of the perspectives, attitudes, and behavior patterns of the target population, which will not be fully captured by other modes of data gathering. Moreover, qualitative interviews can be used to generate hypotheses and propositions which can then be tested on a wider population using a structured questionnaire.

The potential of qualitative interviews has not been fully realized by monitoring and evaluation staff. Most of them, trained in agricultural economics or statistics, are more familiar with quantitative data collection and analysis procedures and have not been exposed to qualitative methodology. Both the staff and project managers tend to be suspicious of information based on qualitative interviews; it is considered "soft," impressionistic, and not able to carry as much weight as numerical estimates derived from structured surveys. Project monitoring and evaluation would benefit if this neglect were corrected. To do so would require monitoring and evaluation staff to develop skills in qualitative interviewing. A conscious effort would also need to be made to encourage decisionmakers to realize the utility of qualitative information. This chapter thus focuses on qualitative interviews of individuals. It presents the general typology of qualitative interviews, guidelines for conducting them, and their limitations. The next chapter describes a particular variant, group interviews.

Types of Qualitative Interviews

Qualitative interviews are distinct from the more familiar structured interviews, which use a formal questionnaire administered by enumerators

trained only to ask preset questions. There are some similarities—in each case a respondent must be approached and his¹ cooperation elicited—but there are essential differences in their nature, the techniques used, the role of the interviewer, and the type of analysis conducted. Qualitative interviews are not limited by a set of predetermined questions to be asked in a given sequence; instead, an interview guide lists the topics to be covered. If certain questions are agreed upon in advance, they are posed in an open-ended way to encourage free conversation and inserted in the interview as the conversation flows. The interviewer enjoys a pivotal role in qualitative interviews; without a formal questionnaire, it is possible to pursue avenues that open as the interview develops. Another difference is that the information generated by qualitative interviews is not usually in the form of numerical or coded data but is contained in notes and summary transcriptions. This is both an advantage and a disadvantage.

Qualitative interviews are usually classified according to three broad types: informal, conversational; topic-focused; and semistructured, open-ended.

Informal, Conversational Interviews

In informal, conversational interviews, the interviewer enjoys complete freedom and flexibility to explore a broad subject with the respondents, who are encouraged to share their views, experiences, values, and information. Issues as they emerge can be further pursued by the interviewer. The interviewer takes few notes during the interview so as to preserve the informal atmosphere.

Despite outward appearances, such interviews must be more than casual conversations. The interviewer has a purpose in mind and must control the conversation to serve that purpose. For example, the monitoring staff of a project talk with several farmers in the project area to gain a better understanding of the farmers' problems in obtaining short-term credit. The talks cover many topics and issues not directly related to credit, but are guided so as to provide insight into farming and local issues that affect the way the farmers perceive their credit needs and the means of satisfying these needs.

This type of interview has several limitations. First, it can be very time-consuming; the conversation can become unfocused and wander around in circles as the respondent expresses views on many issues unrelated to the inquiry. Second, information gathered from one respondent may not be comparable with that from another. When respondents focus on

1. Throughout this book, "he," "his," "manpower," and so on are used generically and are not meant to imply sexual bias.

different issues, it becomes difficult to assess whether or not broad agreement exists on any one of them. This problem is aggravated when several interviewers are involved. Third, this type of interview is highly susceptible to “interviewer effect”: because the setting is informal, the respondent is more prone to be influenced by the personality and views of the interviewer. For example, if the interviewer is particularly warm and friendly, the respondent may reciprocate by avoiding strongly negative statements.

The main strength of this type of interview is that a wide range of issues may emerge—some of which may have been unforeseen by project staff. When the interview is handled with skill, much may be revealed that would not be volunteered in more formal settings. The informal interview is especially useful for diagnostic studies when the problem that project managers see emerging was unforeseen and project staff know no obvious reasons for it.

Topic-Focused Interviews

Topic-focused interviews are conducted using an interview guide which lists the main topics and subtopics to be covered (see example 1). The interviewer, however, exercises discretion in using the guide; he chooses how to phrase questions, asks them in a way that permits a smooth flow of conversation, and dwells in detail on matters which particularly excite the respondent’s interest. Although the guide is used, it does not unduly constrain flexibility in pursuing the conversation.

The interview guide can also provide general instructions to the interviewer, especially when several interviewers conduct a single study. It is important that the initial list of topics and subtopics be limited; otherwise, there will be insufficient time for in-depth and candid discussions. But it should not be so brief that relevant items are excluded. The guide itself can be drawn up on the basis of issues raised in a small preliminary study which uses informal, conversational interviews.

EXAMPLE 1. A Guide for a Topic-Focused Interview

A project is promoting the planting of an improved variety of maize that requires the special application of fertilizer. Monitoring data indicate that the demand for the new seed and fertilizer is not increasing as predicted. The monitoring and evaluation staff are therefore asked to elicit the farmers’ reasons for nonadoption.

The interview guide would list topics and subtopics similar to the following:

- *The farmer’s understanding of the composition of the technical package. (What is it all about? What does it involve? What steps are necessary if one intends to adopt it?)*

- *The farmer's perceptions regarding the advantages and costs of the technical package. (What may be gained by cultivating the new variety? Is it economically profitable? What investments does it require?)*
- *The farmer's judgment concerning the relevance of the package. (Are there constraints on land, credit, or labor that prevent him from adopting it?)*
- *The farmer's views regarding the availability of services. (Does the farmer think he could obtain the needed inputs? What confidence does he have in the available extension advice?)*
- *The farmer's assessment of the risks involved. (Is the new variety seen as less reliable than the traditional variety? Are there consumption, preparation, or palatability problems associated with the new variety? On what basis does he make such an assessment?)*
- *The farmer's assessment of the potential rewards. (Is he interested in increasing production given existing prices?)*
- *Background information. (Availability of labor in the household, size of holding, cropping patterns, and so on.)*

The interviewer will have complete freedom to cover the above topics and subtopics in any order which seems appropriate in a given case. But all the items listed in the interview guide should be covered. More time, however, should be spent discussing issues about which the respondent has particularly emphatic feelings, and those in which the respondent has no interest can be passed by quickly.

If the farmers are classified into those who have never used the new variety, those who tried and rejected it, and those who are currently using it, different guides can be provided for each type.

Topic-focused interviews have at least two advantages over the more informal type. First, since all the interviewers cover the same topics, the data generated are more comparable. Therefore the analysis will be able to indicate the relative priority of various issues: in example 1, the majority of the informants could have indicated that they were not in a position to take the risks involved or suggested that the taste of the new variety of maize was not acceptable to their families. Second, the discussions stay within the context of the subject of interest and thus save time.

Semistructured, Open-ended Interviews

Semistructured, open-ended interviews are the most structured form of qualitative interviewing. They use an open-ended questionnaire which lists the specific questions to be asked.

Semistructured interviews superficially resemble the interviews conducted for structured surveys but differ from them in three main ways. First, the questions are open-ended; respondents are encouraged to ex-

press themselves fully rather than respond to a predetermined list of options. Second, the sequence of the questions is not predetermined; the interviewer is still allowed to exercise discretion in controlling the course of the interview. Third, additional questions can be asked in order to pursue interesting leads.

Semistructured, open-ended interviews have several strengths. First, the information obtained specifically answers certain questions that project managers wish to address. Second, the information from various respondents is comparable enough to determine the simple frequency of responses, although the main emphasis will continue to be placed on the in-depth understanding provided by the respondents. Third, compared with the other types of qualitative interviews, success is less dependent upon the interviewer's interpersonal communication skills and grasp of the subject. Fourth, this type of qualitative interview can be conducted more quickly than the others.

The most serious limitation of this type of survey is that when the interviews are conducted by others than the initiating project officer or survey designer the interviewers tend to confine themselves to the written questions only—they do not pursue promising leads and thus in effect conduct a fully structured interview. The value of the information generated is then too dependent on the quality of the questionnaire.

Combining Types of Qualitative Interviews

The types of qualitative interviews just described can be combined in a single investigation. For example, an informal, conversational interview can be immediately followed by a set of semistructured, open-ended questions.

The following factors and conditions should be considered in the selection of the interview type or mix of various types:

- The nature of the information required. Obviously, if comparable information on a few specified topics is the requirement, the semi-structured, open-ended format should be preferred. When a deeper understanding of the respondent's perspective is called for, other forms are more suitable.
- The skills and expertise of the interviewer. Conversational and topic-focused interviews usually require the interviewer to have a high level of interpersonal communication skills and a good grasp of the subject. These skills are often in short supply and thus have a high opportunity cost. Semistructured interviews require a somewhat lower level of professional skills.
- The background of the respondent. Experience has shown that conversational and topic-focused interviews are preferable for well-

informed respondents capable of clearly articulating their viewpoints. Such informants feel constrained by a formal question-and-answer interview.

- The nature of analysis and presentation that is needed to carry credibility with the decisionmakers.
- Time constraints. Informal, conversational interviews are undoubtedly the most time-consuming; indeed, they are most useful when carried out in several sessions.

Guidelines for Qualitative Interviews

Sociologists, anthropologists, and applied farm study economists have contributed a considerable literature on qualitative interviews. Some general guidelines for these interviews have been developed, although they are not as formal as those for structured surveys. The following are particularly relevant for studies in the rural sector.

Initial Contact

The initial contact is very important for any type of interview. But establishing a basis for easy communication with the respondent is particularly important for a qualitative interview, in which it is hoped that the respondent will frankly state his views, perceptions, and concerns.

The appearance, style, and manner of introduction adopted by the interviewer do make a difference. Dress should be appropriate to the setting—expensive clothes, for example, are out of place on a farm and may emphasize the different status of respondent and interviewer. The interviewer's language should be free from jargon, unnecessary technical terms, and pedantic forms of expression.

The purpose and scope of the inquiry must be carefully explained in a way tailored to the needs of the particular audience. Before government officials and professionals are interviewed, the purpose of the inquiry and relevant specifics can be given in some detail, albeit succinctly. When approaching project participants, however, details may be confusing; a brief statement of the main purpose and scope may suffice—together with such promises of confidentiality as can be offered.

Sequencing of the Questions

It is polite to begin with general conversation; for example, the interviewer could ask about the children (if they are around), cultural events (if they are being planned), or any subject which might interest the respondent. The interviewer in turn should volunteer information about

his own background, family, and experiences to help the informant to place him in a familiar social category. Such preliminary conversation helps overcome any initial reservations of the respondent.

The interview itself should then begin with simple questions that require neither further interpretation nor lengthy recall on the part of the respondent. Questions such as "Are you a member of the farmers' club?," "Do you participate in meetings of the club?," and "How many normally attend a club meeting?" are factual and in most cases uncontroversial.

Care is needed when moving on to more sensitive questions. Consider example 2.

EXAMPLE 2. An Interview Involving Sensitive Questions

An agency responsible for the distribution of fertilizers is performing poorly. Evaluation staff have been requested to conduct in-depth interviews with agency officials in order to learn the causes of the problem. Such an interview might begin:

"I know little about the delivery system that has been set up except what I have read in project documents. So let us start from scratch. When did you really start operating?"

Once the informant has given a brief history—which probably will introduce some of the problems faced by the organization—a noncommittal comment can be made, such as, "Oh, so you have faced a lot of difficulties. Can you be more specific about the problem of _____?" This will keep the discussion going. The informant is likely to introduce more detail. Gradually, the probing can begin, with questions such as, "How do you think this [problem] has affected the performance of your organization?"

Now the core of the interview commences.

As a rule of thumb, only after an informant has described some event or activity should the interviewer proceed to ask for his opinions, feelings, and explanations. For example, only when the informant has described the working of an organization should the interviewer ask for his own evaluation of its performance.

As far as possible, the interviewer should begin with the present and then move to questions about the future or past. Answers about the past suffer from lapses of memory and recall biases. Questions about the future are speculative and cause many respondents to be hesitant.

Wording of Questions

By definition, qualitative interviews require the interviewer to frame questions on the spur of the moment. Three major considerations should be kept in mind.

First and most obvious, the questions must be put in an understandable way. There may be significant variations in the style of language and expressions used by different socioeconomic groups. Questions appropriate for senior officials may require rephrasing when addressed to village chiefs.

Second, questions should be phrased in such a way as to elicit detailed responses. One mistake often made is to ask questions that can be answered by a simple yes or no. Such answers are often eminently suitable for structured surveys but not for qualitative interviews, which are designed to provide deeper meanings, in-depth descriptions, and explanations.² This danger is most evident when underprivileged groups are interviewed. Nervousness and reticence tend to produce answers in simple yes or no terms. From the respondent's point of view, risk of controversy is thus minimized and the interview may be shortened. Inexperienced interviewers try to solve the problem by offering lengthy explanations themselves and then asking the respondent to agree or disagree. Such a practice is highly unsatisfactory and may invalidate the interview, for words are being put into the respondent's mouth. Table 1 gives a few examples of alternative wordings of questions that will produce either a brief or detailed response.

The third consideration is that two or more questions should not be put simultaneously. This tends to confuse the respondent, who does not know which question to answer first and in the process of answering one will fail to answer the other. (We say more on the wording of questions in the context of structured surveys in chapter 5.)

Role Playing

In order to make abstract questions more concrete, and thus facilitate empathy and communication, the respondent can be asked to assume a specific role. For example, "What should be the duties of an agricultural extension worker?" may evoke a more considered response if the question is reworded as: "Suppose you are appointed as an agricultural extension worker; what would you do under the present circumstances?" The rephrased question will enable a farmer or village chief to visualize the role more vividly.

Role playing can help to reduce embarrassment in responding to sensitive questions. Investigators studying deviant behavior often use this technique to obtain information. They do not ask a drug dealer how he secures cocaine. Rather, they ask him: "Suppose somebody wants to procure cocaine. How would he get it in an area like this?" When interview-

2. This point is well elucidated by Michael Q. Patton in *Qualitative Evaluation Methods* (Beverly Hills, Calif.: Sage, 1980), pp. 213-19.

Table 1. Form of Questions Leading to Varying Responses

<i>Leading to yes or no response</i>	<i>Leading to detailed response</i>
Have you heard of the extension services operating here?	What do you know about the extension services operating in this area?
Do you think that if you use _____ your production of maize will increase?	What is your view about the likely result of using _____?
It is sometimes said that the marketing agency only deals with large farmers. Is this true?	What is your view regarding the way the marketing agency operates?
Has the program been successful?	What effects has the program had on you? On your neighbors?
Do you have difficulty in obtaining [specify inputs].	Please describe how you go about obtaining the [specify inputs].
Would you grow more _____ if the government raised the price?	What would be the effect on _____ if the government raised the price?

ing a farmer about crop sales in a situation that has constraints on who may legally buy, the question is better put thus: "Suppose you were a grower who was looking for the best price. How . . . ?" A candid answer may still not be forthcoming, but the chances of it are improved.

A variant of this technique is for the interviewer to assume a role. For example, a question might be put as follows: "Suppose I am the manager of a cooperative store. What would you like me to do?" Or, "What advice would you give me to improve the running of the cooperative store?"

Respondents should be asked to assume only roles with which they can readily identify. It is unwise to ask a farmer, "What would you do if you were the minister of agriculture?" The respondent may regard the question either as a joke or, worse, as holding him up to ridicule.

Probing

Skillful probing is essential in seeking elaborations, details, or clarifications. The success of a qualitative interview largely depends upon the capacity of the interviewer to probe without annoying the respondent. Some have the natural ability to do this successfully, but for many training and experience are necessary. The conversational tone must be maintained; the respondent must not feel cross-examined.

When more details are required on a point raised by the respondent, signals can be given to encourage elaboration. A nod of the head or a simple yes may be sufficient to spur the respondent on. If not, remarks such

as "This is a crucial subject, and I would appreciate it if you would give me more details" or "I am getting the picture, please continue" are usually sufficient to provide the needed stimulus.

It is often necessary to encourage the respondent to be more specific. A reply such as "Oh, farmers like me do not get credit from the cooperative society" can be followed by probes such as "Do you know of a specific case in which someone was denied credit? When did it happen?" Simple questions such as "How did it happen?," "When did it happen?," "Where did it happen?," or "To whom did it happen?" draw the respondent into the desired specificity.

The interviewer should never give the impression that he considers the respondent inarticulate. On the contrary, he must assume blame for failing to understand a comment such as "I am sorry, I missed the point. Will you kindly repeat it?" or he must rephrase the question to stimulate a revised answer. If a clear answer is not obtained after a second attempt, the question should not be pursued. Perhaps the respondent has nothing to say or does not want to say it. Persistence can be dysfunctional in such a case. If the question is important, a further attempt may be justified later in the interview.

Controlling Conversations

Interviewers encounter situations in which the respondent gives long and seemingly irrelevant answers. For example, a question on the effectiveness of a marketing agency produces a discourse on corruption in public institutions. It is important that the interviewer seek to understand what the respondent is trying to communicate here; indirect answers may convey pertinent ideas and issues which the interviewer has overlooked. The respondent who mentions corruption is perhaps hinting that financial irregularities exist in the marketing agency but is careful to avoid saying it openly. Points which appear to be marginal or irrelevant at first can come to be regarded as significant and relevant in later analysis.

Some conversations, however, go seriously off the point and if not controlled will waste time. One step is to cease giving the respondent cues to continue talking. Stop nodding the head, saying yes, or taking notes. Glance away, breaking eye contact. Most respondents are sensitive to such nonverbal reactions, but if such attempts are unsuccessful interrupt with statements such as "What you said is very enlightening and I understand your point; now I would like to know. . ." or "What you have said prompts me to ask another question." We stress the obvious point that the interruption must be made without giving offense.

Neutral Attitude

An interviewer should be both a sympathetic listener and a neutral observer; he should avoid giving the impression of having strong views on

the subject under discussion. This neutral attitude should not be a mere facade; the respondent is entitled to his own views—the interviewer's role lies in eliciting them. Despite occasional temptations, attempts to convert the respondent to a point of view must be avoided.

Several strategies can be pursued for dealing with controversial issues. One is to stress that the purpose is to search for information and that the interviewer will be able to form an accurate judgment only when the study is over. Such a posture, though evasive, can reassure the respondent. It may of course require that the interviewer try to distance himself from specific activities of the project to which he is formally attached.

A second strategy, known as the illustrative example format, consists of stating both sides of the issue to demonstrate familiarity with it without taking sides. For example, consider a series of in-depth interviews with farmers who have not repaid their loans to the agricultural credit bank. Besides assuring them that the interviewer has not come to collect the loans or to persuade them to repay the balance, the interviewer can also state an appreciation of both sides of the issue. On the one hand, the bank is right when it insists that those who borrowed should strictly follow the terms of the contract. On the other hand, the interviewer can sympathize with borrowers who are not in a position to repay. Therefore, he has come to understand the problems which have placed them in this unfortunate position.

A third strategy is to candidly express one's views and engage in an honest dialog. This strategy is rewarding only when the informant is knowledgeable and does not feel constrained in expressing his opinions. Although an interviewer can follow this course with government officials, experts, or project staff, he might not find it suitable in interviewing small farmers, who are usually reluctant to say anything which can contradict the visitor in any way.

Recording the Interview

The advantages of a well-conducted interview are lost if an adequate record is not made either during or immediately after the interview (especially since the informal conversational interview is not conducive to note taking during the conversation.)

The use of a tape recorder releases the interviewer from elaborate note taking, but consideration is needed before using one with respondents who are unaccustomed to it as it may lead to an unnatural response. Even when a recorder is used, there are several reasons why notes should be written up. First, the nonverbal behavior of the informant may be relevant. For example, if the respondent becomes excited when discussing the working of a participatory organization, this should be noted, for it will be relevant when interpreting the interview data. Second, the inter-

viewer needs to note his own thoughts which are stimulated by the replies of the informant. Third, in the event of a malfunction of the tape recorder, the interviewer can fall back on his notes.

When notes are taken during the interview, the questions need be noted only by a code or number if an interview guide is used. The interviewer should use quotation marks whenever he reproduces the language of the informant. Quotes are helpful for writing reports and making verbal presentations to decisionmakers, but it is important that they be accurate, and they should be used very sparingly. As far as possible, the interviewer should develop a system for noting down his own ideas, responses, and feelings. Fresh ideas and insights stimulated by the respondent's replies may be lost unless noted at the time. It is advisable to write such comments in brackets in order to distinguish the interviewer's ideas from those of the respondent.

There is a universal agreement among practitioners that the interview should be written up as soon as possible—ideally, immediately following the interview. Recall lapses are minimized in this way. Multiple interviews in one location should be spaced to allow the notes to be transcribed after each one. Transcribing interviews from tapes is very tedious and time-consuming. In any case, the real feel of an interview cannot be captured without substantial editing and adding the interviewer's reactions and impressions. We recommend that the interviewer listen to the tape, supplement his notes on the basis of the verbatim recording, and then prepare the summary of the interview, including a description of the setting and the respondent's nonverbal behavior, credibility, and knowledgeability.

Reliability of the Interview

How is the reliability of the information generated by a qualitative interview to be assessed? How can we be sure that the respondent has provided accurate information? This problem, of course, is not unique to qualitative interviewing; it is common in all types of interviews. But because of the subjective nature of the written summary, the issue of reliability is particularly pertinent in this context. Because by definition there is no totally objective test that can be applied in qualitative interview situations, judgment of reliability must be based on an assessment of respondent-related factors. Some are described below.

Knowledge

Obviously, the first consideration is the knowledge that the respondent may be expected to have (a particularly important point when interviewing key informants). Questions for a checklist include:

- Is the respondent's knowledge of the matter direct and firsthand?
- Is the respondent in a position to provide accurate information?
- If the respondent is relying on secondhand sources, are these sources credible?

A respondent may be knowledgeable about some items and relatively ignorant about others. For example, one trader may provide all pertinent information about retail practices but not about the wholesale trade. With another the reverse may be true. Therefore the interviewer should ask himself the above questions with reference to each of the principal subtopics in the interview.

Credibility

Some people have a tendency to boast; others have a fertile imagination and unconsciously exaggerate; still others aim to enhance their self-importance by giving misleading answers: the interviewer should therefore assess the respondent's credibility. Questions for a checklist include:

- Is the respondent eager to make strongly authoritative statements?
- Does the respondent consider before replying and seem perceptive about the issues?
- Are the respondent's answers based on practical considerations?

Ability and Willingness to Respond

Some respondents find it difficult to articulate their feelings, judgments, and opinions to outsiders. This problem is compounded when the interviewer comes from a higher socioeconomic stratum.

A slightly different problem is encountered with qualitative interviews of senior officials who are pressed for time. Although they have the ability to express themselves, they have little time to go into details. Thus the interviewer ends up with an incomplete picture, which can bias the findings.

Ulterior Motives

Respondents may have an ulterior motive for providing inaccurate information. Extension staff may exaggerate the performance and impact of agricultural extension services. A health worker may magnify the problems encountered in reaching out to target populations. Staff directly involved in project efforts have a professional stake in promoting their ac-

tivities and covering their shortcomings; often this bias is more subconscious than a deliberate attempt to mislead. Questions for a checklist include:

- Was the respondent trying to paint a positive picture?
- Was the respondent trying to rationalize a distasteful fact?
- Was the respondent dwelling excessively on problems and difficulties in order to seek sympathy?

Bars to Spontaneity

The social context of the interview also affects the expression of ideas and opinions by the respondents. For example, when a farmer is interviewed in the presence of government officials or project staff, he might not reveal the truth because he is afraid to antagonize them.

Questions for a checklist on this include:

- Were there some people whose presence might have affected the respondent's answers?
- Was he anxious that others might overhear him?
- Was the location private enough to ensure total confidentiality for the interview?

Desire to Please

There is a tendency for respondents to give answers which they believe the interviewer desires either from politeness or in the hope of shortening the questioning. In such a case it is particularly important to avoid giving the respondent clues regarding the interviewer's opinions. Questions for a checklist include:

- Did the respondent show undue deference?
- Did the respondent seek the interviewer's opinion before replying?
- Did the interviewer say anything which silenced the respondent or changed the thrust of his responses?

Other Factors

Finally, one should not forget that recent events might have influenced the views expressed by the informant. Consider the case of a farmer whose application for short-term loans has been turned down recently by the project. Obviously his disappointment may be reflected in his assessment of the project. The mental and physical states of the respondent

also affect his responses. When he is tired, he can be irritable and react negatively to questions. When he is in a good mood, he is likely to be more patient with the interviewer.

General Respondents and Key Informants

Qualitative individual interviews can be conducted with both general and key informants. The simple difference between them is that general respondents primarily give information about themselves, whereas key informants provide information about others or specific situations or conditions existing in the area. The key informants are essentially knowledgeable individuals who are in a position to provide relevant information, ideas, and insights on a particular subject.

Experience has shown that village chiefs, teachers, traders, officials of cooperatives, and local government officials make good key informants in rural areas. They understand local customs, traditions, and social and economic conditions and are able to express themselves well. At least some of them can take an objective view of the issues. And even those with a vested interest may prove to be useful if their allegiance and prejudices are taken into account in analyzing their observations.

Obviously, different studies necessitate key informants of different backgrounds and experience recruited from varying occupational groups, socioeconomic strata, and organizations. For example, for a study of an agricultural extension service, extension staff, knowledgeable farmers, agricultural specialists, and local government officials will make excellent key informants.

Key informants should be carefully selected to reflect diverse viewpoints and concerns. The ideal method is to identify various sources from which the key informants for a study can be drawn and select a few from each of them. If it is revealed during the course of the interviews that there are persons who possess highly relevant information and ideas, they can be added to the list of informants. Usually a key informant study provides for interviews of fifteen to twenty-five respondents.

Limitations of Qualitative Interviews

We have referred to specific weaknesses of particular types of qualitative interviews. But three general limitations of all qualitative interviews must be borne in mind.

First, they do not generate quantitative data that can be summarized to provide valid general estimates. Because the responses are so varied in content and context (which is the strength of the method), it is difficult to summarize the results in order to say, for example, that 60 percent of the

farmers are satisfied with the existing extension services or with the credit institutions.

Second, it is rare for in-depth qualitative interviews to be used with probability samples. By definition, key informants are a biased selection from the general population. One common source of error in such interviews is that the investigators tend to have an elitist orientation and select their informants on the basis of their social and economic status rather than knowledge and experience. For example, it is not uncommon to rely largely on village elites for understanding the problems of smallholders or on government officials for understanding the problem of nonutilization of credit by farmers.

Third, the findings are susceptible to biases which arise out of the inaccurate or distorted judgments of the interviewers that result from their shortcomings in cognitive processing. For example: the interviewer tends to pick up information and ideas that confirm his preconceived notions; he gives more importance to the views of the elites than to those informants of low socioeconomic status; and vivid descriptions and selective data leave a greater imprint on him than do abstract ideas and explanations.

Nevertheless, as we stated at the beginning of this chapter, qualitative interviews are underused in most monitoring and evaluation systems. When insight is needed on reasons for unexpected reactions by the target population, such interviews can be extremely helpful.

3

Conducting Group Interviews

THE BASIC PRINCIPLES of qualitative interviews were introduced in the previous chapter in the context of interviewing individuals, whether general members of the population or key informants. Qualitative interviews can also be conducted with groups acting together and from individuals making up a group. Such interviews can be conducted by one or more interviewers, with or without an interview guide, and with groups of varying size and composition. Superficially, the difference between individual and group interviews is the number of respondents participating in a single interview session; however, this basic difference leads to major variations between them with regard to planning, the nature of interview guides, probing techniques, and the control of discussions. We discuss these variations in this chapter.¹

There are two types of group interviews, community interviews (CIs), to which all members of a community or village are invited, and focused group interviews (FGIs), which are limited to a few selected individuals. CIs are widely used by researchers in agriculture. They are conducted on the basis of an interview guide and take the form of public meetings, usually called at short notice, with a large group (more than fifteen persons), which limits the opportunity for discussion among all those present. Only a small number of questions is asked, and each participant is not expected to answer all the questions individually. An interdisciplinary team rather than a single interviewer is more effective in this situation.

An FGI, conversely, is conducted with a small group (six to ten participants). One of its distinguishing features is that the participants discuss ideas, issues, insights, and experiences among themselves. Each member is free to comment, criticize, or elaborate on the views expressed by the previous speakers. The moderator (the word "moderator" is more appropriate than "interviewer" in this case) guides the discussions using vari-

1. This chapter draws on a monograph by one of the authors, Krishna Kumar, entitled *Conducting Group Interviews in Developing Countries* (Washington, D.C.: U.S. Agency for International Development, 1987).

ous probing techniques. The participants are selected on the basis of criteria, which vary depending on the objective of the inquiry (discussed below).

There are three main reasons why group interviews may be preferable to individual interviews for project monitoring and evaluation. First, group interviews enable the investigator to gather information in a rapid and economical manner. One can interview eight to ten persons in an hour or two, whereas to interview them individually might take two or three days. Although a group interview will not provide the same depth of information that can be gained from individual interviews, a skilled interviewer can obtain considerable information and understanding. Although only a very small team is required, rather than a large enumerator force, the team must possess interpersonal skills and often a range of professional skills. The full collaboration of project staff may be needed to make up this team.

Second, group participation sometimes reduces individual inhibitions and thereby provides information which might not otherwise be revealed. People often are willing to share feelings, emotions, and concerns in groups which they are reluctant to do in more private settings. They find a sense of security in the group, which is undoubtedly an important consideration in rural areas where respondents suffer from a feeling of inferiority in the presence of outsiders. If other farmers in a group state their reservations about a recommended technical package, a cautious farmer may be led to express his own doubts. Sometimes, however, the opposite effect occurs—the individual may not wish to express his views in a more public setting.

Third, information gathered in group interviews is sometimes more accurate than that obtained in individual interviews because respondents are reluctant to give inaccurate answers when they may be contradicted by other participants. Or, if they do, they may well be corrected. In one well-known example of group interviews in India, a large-scale landowner said that he owned only 30 acres of land—the maximum permitted under law—but grudgingly conceded that he owned 300 acres when the other participants humorously questioned him.²

Community Interviews

The disappointing fact about community interviews is that even when they are conducted little if any mention is made of the information gained from them in project reviews and evaluations. The reason often lies in the way they are conducted. Most of the meetings are not carefully

2. Wolf Ladejinsky, "The Green Revolution in Bihar: The Kosi Area." *Economic and Political Weekly* 4, no. 39 (September 27, 1969).

planned, and structured interview guides are not followed. Interviewers generally fail to involve the majority of participants in the process and do not follow suitable probing techniques. Moreover, the discussions are not systematically recorded; interviewers largely rely on limited notes and memory. Under these conditions, it is not difficult to see that investigators find it prudent not to acknowledge the source of their information, for serious questions can be raised about the validity and reliability of CIs conducted by them. We believe that CIs can generate useful, reliable information if the few simple steps outlined below are followed.

Structured Interview Guides

Some practitioners believe that CIs should not be conducted using a structured interview guide which lists specific questions because it stifles the creativity of the interviewer and reduces the interest of the participants by interfering with the flow of discussion. There are advantages, however, in constructing structured guides. As discussed in the previous chapter, such guides assist in the collection of comparable, systematic information and help to keep the discussions focused. One problem with community interviews in the absence of a guide is that the discussions tend to drift and interviewers get carried away by interesting, but irrelevant, topics.

Also, without the help of a guide, too much reliance is placed on the ability of the interviewer to moderate the discussion and pose questions in an understandable form. The difficulty is further compounded when interviewers come from other cultures than the interviewees. In such situations, a structured guide is a great asset because interviewers can rely on preformulated questions.

The guide should not be used to enforce a rigid course of action. Only a limited number of questions should be included in it, usually not more than fifteen. Some important rules for framing the questions are:

- The language should be simple; technical jargon and folksy expressions should be avoided. The questions should be understandable by the least-informed member of the community at the meeting.
- Politically or culturally sensitive questions should not be used. Societies and communities have their own taboos, inhibitions, and sensibilities, which should be respected. Many questions that can be asked in individual interviews cannot be raised in community meetings.
- Questions on large, controversial issues should not be included. They can generate strong emotions that contribute to overt conflicts and tensions in a meeting.

Selecting Communities

Conducting interviews in all or even most of the communities covered by a project or program is rarely feasible because of the constraints of time and resources. Hence, a few communities should be selected carefully to represent the general populations of interest.

In most cases, the investigator cannot use probability sampling to select the communities. Rather, he relies on two informal sampling techniques. One technique involves classifying communities according to objectively verifiable criteria and selecting a number from each of them. For example, for a study of the effectiveness of extension services, villages can be categorized on the basis of their size, existence of a shop selling inputs, and accessibility by an all-weather road, and then one or two villages can be selected from each category for conducting community interviews. The second widely used technique is to select communities on the basis of expert advice. In the above case, the chief of the agricultural extension services can be asked to identify the villages in which extension services have been effective and those in which they have not. However, to guard against bias (the extension chief may want to present only a positive picture of the extension services and may deliberately misguide the investigator) or a misinformed source, more than one expert should be consulted.

Size and Timing of the Meeting

The interviewers have little control over the size of a community meeting. Size depends not only on the population of the community but also on such factors as time of day, convenience, adequate notice, and local interest. There will always be uncertainty about the number who will participate, but experience shows that if the meetings are well-publicized the attendance will be large, even if curiosity is the main motivation.

Obviously, the CIs should not be held at a time of day when people are involved in their work. Unfortunately, this advice is not always followed in remote villages which are difficult to reach by outsiders. The interviewers often schedule a meeting in late morning or afternoon, which is certainly not a convenient time for farmers. The result is that participants are not representative of the community or are forced to attend against their will. In either case, the reliability of the information and recommendations generated become questionable. The prime consideration should be the convenience of the participants and not of the interviewer. If the size of the group exceeds thirty, consideration should be given to dividing into smaller groups.

Interview Team and Protocols

Although CIs can be conducted by one interviewer, a team is preferable for three reasons. First, if the group is large it is extremely taxing for one interviewer to preside over the meeting, ask penetrating questions, probe the respondents, and also take extensive notes. Second, the value of the notes improves if several independent sets are taken and compared. Third, team members, especially when they have different disciplinary backgrounds, complement each other in probing respondents, which improves the quality and depth of the information.

A few ground rules should be followed by the team so that the CI proceeds smoothly. Team members should coordinate their interventions, making sure that the participants get a fair chance to respond on each topic introduced by each team member. There should be prior agreement on how a team member should pick up the probing role as another completes a probe on a specific point. Of course it is not necessary that each team member ask questions on every topic. Interviewers should refrain from interrupting each other. Even though a question springs to mind it should be noted and brought up at the appropriate moment.

The temptation to help a colleague by trying to interpret a participant's response should be resisted unless the other team member is really misunderstanding the answer and thereby creating an embarrassing situation. And interviewers should avoid undue repetition of questions which other team members have already posed.

Balancing Participation

One of the most difficult tasks for interviewers in CIs is to restrain a few community elites from monopolizing meetings. Such prominent individuals believe they speak on behalf of the others, who the elites consider unable to express themselves. These leaders have their own interests to promote and may not be reflecting the true views of the other participants. The result is that the very purpose of CIs is defeated, and the result is not community interviews but interviews with a few elites which would have been better conducted individually.

Restraining the leaders requires consummate acumen and interpersonal skills on the part of the interviewer. One strategy that has proved effective is to meet with the leaders before the interview and seek their views on some of the subtopics to be covered. This strategy has two merits. First, the leaders will be less likely to repeat themselves in the large meeting since they have been able to articulate their views and concerns in private. Second, the interviewer has an excuse to say publicly that he or she has discussed the subject with several leaders and has now come to hear other members of the community.

All community members at the meeting cannot be expected to participate in discussions, but every effort should be made to secure the partici-

pation of the majority. The interviewer can seek to balance participation in two ways. First, he can specifically address questions to reticent persons. For this purpose, he can look in their direction and say, "I would very much like to hear from you. What have you to say about it?" Second, he can take polls on selected questions by asking participants to give their responses by raising their hands. Example 3 shows how an interviewer was able to encourage participation by his humorous remarks.

EXAMPLE 3. *Involving Participants in Community Interviews*

A series of community interviews were being conducted in an East African country for evaluating an area development project that had been extremely successful in motivating farmers to establish farmers clubs. The team leader, concerned about the domination of meetings by a few leaders and by the nonparticipation of women, included the following remarks in his introduction.

When I was coming here, my boss called me and told me that he was interested in knowing the views of all the people in a meeting; he could not make his decisions on the basis of the opinions of a few individuals. In fact, to tell you the truth, he promised me a raise if I succeeded; otherwise he might even fire me. So please promise me that all of you will participate in discussions. If you don't, you will have to give me a piece of land so that I can join you [laughter].

The team leader then turned to the women, who usually sat separately, and added:

But this is not the only problem that I have. My wife has heard a lot about you and your participation in farmers clubs. She wants to know more about what you have been doing, what your experiences have been, and what can be done to improve your participation. If my boss gets angry he can only fire me, but if my wife gets upset I might be in greater trouble.

These remarks gave the team leader an excuse to humorously probe the participants in the meetings. Whenever some people were not participating in the discussions, he would simply say, "Oh, my friends, you seem to be forgetting my problem." Group members would laugh and respond to his questions.

Source: International Fund for Agricultural Development.

Generating Community Data

CIs can (with certain limitations) generate aggregate and general data at the community level. Participants, for example, can give approximately the number of families in the village, available health services, the number of houses which have direct access to water, the number of local resi-

dents who work in nearby towns, and the number of children attending school. They can also provide some aggregate data on farm inputs and outputs, extent of mechanization, distances to markets, and even cropping pattern and production levels. Example 4 gives some of the questions used in a project with a large number of group interviews.

EXAMPLE 4. Group Interviews for Generating Quantitative Data in Costa Rica

In a research project in Costa Rica, a structured questionnaire was used for group interviewing in 860 communities for generating community-level statistics. Below are some examples of the questions.

(7) What is the daily wage of an agricultural worker in this area? _____

7-1 For how many hours _____

7-2 Does this include

	Yes	No
Food	_____	_____
Housing	_____	_____
Land for growing food	_____	_____

• • •

(9) What are the three main crops here? _____

9-1 Which is the most important? _____

the second most important? _____

the third most important? _____

9-2 How much is sold commercially?

	Almost all	More than half	Less than half	little
Crop				
1				
2				
3				

• • •

(14) Where do people generally go to buy the things they cannot buy here?

Community _____

District _____

County _____

14.1 How do they go there? _____

14.2 How long does it take? _____

• • •

(37) Is there a high school here? Yes _____

No _____

(38) Is there a grade school here? Yes _____

No _____

The authors complemented their information with direct participant observation and other sources of data. It should be mentioned here that in this study group meetings comprised selected community leaders and key informants.

*Source: Jeffery Ashe, *Assessing Rural Needs: A Manual for Practitioners* (Washington, D.C.: Volunteers in Technical Assistance, 1978).*

When such data are gathered through CIs, the interviewer should encourage different participants to verify the information. If one speaker says that thirty persons from the village work in the neighboring town, the interviewer can point to a few participants and ask, "Do you agree with this estimate?" The interviewer should also try to understand the basis on which respondents have made their estimates—in some instances they will know the individuals referred to by name.

Opinions on certain subtopics can be ascertained by polling the participants. In such cases, the questions should be able to be answered with a simple yes or no. The interviewer asks a question and requests participants to answer it by raising their hands. The results of such polls should, however, be treated cautiously; there is a real danger of superficiality in such a technique.

The quantitative data generated through community interviews can be aggregated and analyzed in two ways. First, individual respondents can be treated as cases. For example, suppose ten group meetings are attended by a total of 200 farmers. If 80 farmers say that they received credit for the purchase of oxen, it can be reported that 40 percent of the farmers interviewed received credit. Second, each group can be treated as a case. In the above example, the investigation might report that in only three villages out of ten did the majority of farmers receive credit.

One should be extremely careful in evaluating aggregate findings generated by CIs. They have some validity only if three essential conditions are met. First, participants must be representative of the target populations. In the case of CIs for ascertaining support for community centers, if the meetings were held at noon when most of the adults were working on their farms, the findings would give a distorted picture to the extent the decisionmakers would not be represented adequately in them. Second, group processes must not inhibit free expression of views or preferences. For example, many respondents might not like to acknowledge in public that they will not contribute toward the construction on a community center in their village. Third, the questions must not be politically or culturally sensitive. If sensitive questions are asked, participants are unlikely to provide candid answers.

Postmeeting Conversations

Providing opportunity and time for individual conversations after the meeting should be regarded as an integral part of the process. Some shy

participants prefer to bide their time and approach the interviewer at the end of the meeting. Others might regard it as impolite to contradict previous speakers during the meeting. Still others may have failed to catch the eye of the interviewer at the time.

Before terminating a meeting, the interviewer should indicate that there is time to discuss any relevant issue after the meeting with individuals who wish to stay on. In a CI session in a South Asian country, the interviewer had the impression that certain people were eager to speak but did not. At the end of the meeting, he engaged them in conversation and found that they were in fact the smallholders, who felt that the village leaders were misusing project resources to further their own interests. Further inquiries revealed that they were right. Postmeeting conversations can be extremely rewarding, and time (up to an hour) should be allowed for them.

Focused Group Interviews

The procedures for focused group interviews are similar to those for community interviews, but there are some differences because of the smaller number of participants, the fact that the participants have been selected according to some contribution they are expected to make, and the stress on spontaneous interaction between the participants.

Interview Guide

The only type of interview guide that is recommended for FGIs is a short checklist to remind the moderator of the main subtopics to be covered. It is thus much less structured than guides used in CIs and does not detail the questions nor provide detailed instructions to the moderator.

One of the primary objectives of FGIs is to explore subtopics in depth by interviewing a small, carefully selected group. Therefore the number of subtopics should be limited. Moreover, the few subtopics are likely to generate discussions that raise issues not anticipated by the investigator. In an FGI on the use of a specific variety of fertilizer, for example, one of the farmers may mention that although smallholders would like to use them, they encounter problems in bringing them from the town because of the long distance they have to walk. This casual remark can lead to an interesting discussion of the mechanisms for the distribution of fertilizers. Participants are likely to propose different ideas and suggestions which can be further examined by the group.

Size and Composition of the Group

The optimal number of participants in an FGI can range from six to ten. Within this range smooth conversation is possible and the moderator can

steer the discussion without depriving individuals of a chance to air their views.

Smaller groups—three or four persons—make participants feel pressured to constantly comment whether or not they have something relevant to say. Moreover, the discussions are vulnerable to domination by one influential participant. Large groups—more than twelve persons—leave little time for individuals to respond spontaneously or to have a chance to fully present their points of view. Participants often lose interest entirely or form subgroups which engage in fragmented conversations outside the moderator's control.

It may be necessary for the focused group to be homogeneous in social composition. In the stratified societies of the developing world, participants drawn from different social and economic strata may be unwilling to interact on the basis of equality. Status differences impinge on interpersonal communication. Persons of lower status generally are reluctant to talk in the presence of their perceived superiors.

Members of a group usually should not know each other, for anonymity minimizes status barriers. This requirement often cannot be met, however, in a rural community. People tend to know about each other even if they are not acquaintances.

Selection of the Participants

The best approach to identifying participants is to consult key informants who are knowledgeable about local conditions. It is always prudent to consult several informants to minimize the biases arising out of individual preferences. Once the list is prepared, the investigator can select the required number from it.

Efforts should be made to include diverse participants. One can achieve this by classifying the target population on the basis of carefully selected criteria relevant to the study and then selecting participants from each category.

This can be explained with a simple illustration. Consider the owners of tractors in an area development project. They can be classified on the basis of variables such as age (young or old), gender, size of holdings (small or large holder), and literacy. They can also be categorized according to the use of tractors (those who use tractors for their own farms and those who hire them to others) and source of financing (those who got loans from the project and those who raised the credit elsewhere). If cost-effectiveness is the topic of debate, the participants should include a range based on size of landholding, source of financing, and the mechanical skills of the owners—all factors that affect the cost-effectiveness of tractor operations. By contrast, if the purpose of the focused group interview is to identify gender-based differences in the use of tractors by

farmers, such elaborate categories are unnecessary—a simple classification based on gender may be sufficient.

Location, Seating Arrangements, and Duration

FGIs can be conducted in any place where eight to ten persons can be comfortably seated and assured of some privacy. The most easily available sites in rural areas are primary school buildings, health centers, and community centers.

Seating arrangements should facilitate maximum interaction among participants. The best course is to have a table around which all the participants can face each other, or alternatively to arrange chairs in a semicircle. If group members prefer to sit on the floor, they can do so. The important thing is that all participants be physically and psychologically comfortable. If possible, FGIs should not be held in the open, where intrusions from outsiders can scarcely be avoided. An FGI can be an interesting social event in a remote village; people whose curiosity has been aroused want to know what is going on, and village officials may insist on participating.

The duration of an FGI should not exceed two hours unless the discussions are of such interest that the group wishes to continue. If only one or two participants wish to continue, the moderator can arrange for post-meeting conversations, as in the case of CIs.

Opening of the Interview

The moderator should explain the purpose and scope of inquiry after the group members introduce themselves. A brief explanation should serve the purpose. Persons recruited for an FGI may have little idea of what is expected of them. It is therefore important that the moderator make the following four points. First, an FGI is not a question and answer session. It is an informal discussion, and a point made by one can be commented on by others. Second, the group is organized to hear the views and experiences of all the participants. Third, the moderator is interested in the entire range of ideas and explanations; if one has a divergent view on any item, he should express it freely. Fourth, because of the constraints of time, each should be as succinct as possible. Example 5 shows how an appropriate introduction to an FGI may be done.

EXAMPLE 5. *Introductory Remarks for a Focused Group Interview*

My colleague and I are grateful that you were kind enough to come to help us. Let me mention the purpose of our meeting here. As you probably know, this project in your area is being funded by AID to try to popularize an improved variety of maize, which should significantly increase yields. The proj-

ect is also providing interested farmers with seed, fertilizers, insecticides, and pesticides, which are necessary for the cultivation of the improved variety. However, most of the farmers are not using it. Why? We have some ideas and explanations for this, but we are not sure about them and so we want to hear your views. I stress that we want to know your real views; the best help you can give us is to be candid. Why are so few farmers using the new maize variety? Is the improved variety profitable to the farmers? Do the farmers experience difficulty procuring the needed inputs? Do they need credit? What are the general impressions about the improved variety? Do the people like the taste of the new variety of maize?

I have a few requests. Because all of us must participate in the discussions, we must be brief and to the point. Please remember that each of us can make comments or raise questions about what others say. This is an informal discussion among friends. So do not hold back any ideas or information. Even if you disagree with the rest, state your views.

We will be taking notes so that we can remember your comments. (We would also like to have your permission to record the discussions on a tape recorder.)

Before opening the main discussion, the moderator should establish some rapport with and among participants by using a warm-up period to chat about general matters. He can exchange a few words with farmers about crop prospects during the current agricultural season. The moderator can ask participants to explain the significance of any impending social or cultural event, such as a fair. The moderator should not allow detailed discussion on such generalities; the purpose is simply to break the ice. Too immediate a formal start prevents the group from settling down to a comfortable, relaxed discussion. During the warm-up period, the interviewer should identify the participants who are reticent as well as those who love to talk and should make a mental note of which ones will need encouragement to express themselves or gently need to be controlled.

Slides, films, or pictures can be shown to stimulate discussion on a specific subtopic. The moderator can, for example, show a documentary about a farmer planting a new variety of maize and then ask the group members to give their views about the actual operations involved, and the advantages and limitations of the maize variety.

Probing and Pacing of Questions

The probing strategies used in individual interviews can also be employed in FGIs. In group sessions, however, the moderator should adopt a posture of "sophisticated naïvete." He should convey the impression that although he understands the subject, he does not know the details as they do. Such an approach usually works because people are generally willing to help. Thus the moderator can ask specific details, saying, for

example, "You know that I am not a farmer, so you will have to explain it to me in greater detail," or "I wish I knew more about the sources of credit in the community. Will you kindly tell us about it?" Such probing induces participants to think more deeply on the subject and to verbalize their feelings and thoughts.

The moderator should seek specifics. Experience shows that if the moderator asks specific questions, others follow his example, making his task a little easier.

To cover several subtopics within the stipulated time is obviously not an easy task given the vagaries of group discussions. Some items may be more interesting than others to the participants but less relevant to the objective of the inquiry. For example, in a focused group discussion on the adaptation of a new variety of wheat seed in an area development project, the young, educated participants might like to dwell on the resistance that they encounter from their elders, but the moderator would like to know whether the necessary inputs are available at convenient locations, whether they are being used, and what suggestions the participants may have for promoting them. It is important that the moderator budgets time for each topic.

Controlling Discussions

One common problem faced in FGIs, as in CI, is that a few articulate persons tend to dominate discussions. They have an opinion on every subject, and some go to an unusual length to make a superficial and often irrelevant point. The moderator has to be careful in dealing with them. Any attempt to interrupt their long-winded remarks might be construed as an offense undermining the spontaneity of the discussion. The three strategies used in qualitative interviews discussed in the previous chapter, also are relevant in this context.

The first is to give nonverbal cues to the participant that the moderator expects him to stop. He can look away or stop taking notes. The second is for the moderator to intervene, saying, for example, that he would like to summarize what has been said so that the group does not misunderstand or misinterpret the comments. The moderator can then refocus the discussion. Third, as soon as there is a pause, he can say, "You have made many useful points which need to be pursued in more depth in another group. I would, however, like to explore another item with this group." This leaves the participant no option but to stop.

Controlling Group Pressures

In FGIs the moderator must be able to minimize group pressure, which inhibits the dissenting participants from expressing their views or encour-

ages them to agree to positions to which they do not subscribe. The underlying reasons for group pressure are complex and varied. Sometimes the idea or explanation proposed is new and most of the participants are momentarily captivated by it. In other cases, the majority has little to add and gets out of this uncomfortable situation by fully endorsing the position put forward. As stated earlier, there may also be the dominance of the articulate and influential.

The moderator can minimize group pressure by encouraging participants to express diverse views and perspectives. As soon as he sees that an idea is catching on without sufficient examination of the alternative positions, he should try to hold back a premature conclusion. He can ask for other ideas, explanations, or recommendations. For instance, if an FGI discussing farmers' views on different modes of delivering fertilizers is taken by the notion that village cooperatives should be the only mechanism for their distribution, the moderator should interject, "But what about other methods? Can we suggest alternatives?" This will lead the group to at least consider other approaches. Or the moderator can put forward an alternative. In the above case, the moderator can say, "What about the local store? Why can't fertilizer be made available there?" The problem with this strategy is that the participants may get the impression that the investigator is pressing for a particular option and may reverse themselves in order to please him. It is therefore necessary to stress that the idea is mentioned not to encourage them to accept it but to generate discussion on the subject.

Finally the participants who seem skeptical to the position taken by the group can be encouraged to express their views. When conducting a discussion, one often gets the feeling that some members are not convinced of a particular position, and yet they remain silent. Often what they need is encouragement by the moderator. In such situations, the moderator can look at one of the reluctant participants and say, "What about you? You might have a different view?" Such a remark can reassure him and he might present his views on the subject.

Recording of Discussions

As with all qualitative interviewing, the ultimate value of an FGI depends on the nature and the quality of the record made. In developed-country market research, video recorders are being increasingly used for filming the discussion, but this is generally out of the question in rural areas. The use of a tape recorder should be seriously considered. With due regard to the cautionary comments in the previous chapter, a selected small group may not be unduly disturbed by a recorder.

Another possibility that is more likely to be practical with an FGI as compared with a CI is to use a rapporteur to record the discussions. It

eases the burden on the moderator and enables him to concentrate on conducting the discussions. Such a role fits well in the format because it is known by most to be standard practice in more formal meetings.

Limitations

Group interviews have well-recognized limitations. Many topics cannot be examined in them. Many people are reluctant to share their views on sensitive issues in public. Individuals tend to dominate the discussions. Moreover, group interviews are highly susceptible to interviewers' biases, which can undermine the reliability and validity of the responses. Three reasons for this can be mentioned. First, interviewers tend to pick up information and ideas that confirm their preconceived notions and hypotheses—they hear what they want to hear and ignore what they want to ignore. Second, there is a normal human search for coherence in disparate, irreconcilable remarks of various respondents; this quest for consistency can lead to oversimplifying the complex reality by overlooking evidence that is not consistent with the findings of earlier interviews. Third, there is a tendency for interviewers to give more credence to views expressed by elites than by others. We take the view that both the quality and credibility of the findings can be improved if proper procedures are followed and if the information generated is cross-checked with that gathered through other means.

4 | Participant Observation

PARTICIPANT OBSERVATION is a type of qualitative data-gathering method that requires direct observation of an activity, behavior, relationship, phenomenon, network, or process in the field. This observation is supplemented both by information gathered through qualitative interviews with key informants and by data from analysis of documents, records, and other sources. The participant observer seeks to go beyond outward appearances and probe the perceptions, motives, beliefs, values, and attitudes of the people involved.

The central concept in participant observation is that the investigator participates in the social reality experienced by the community under observation. For instance, in studying the economic behavior of small farmers, the investigator becomes a part of the rural community to the degree required in order to understand the farmers' perceptions of the constraints and opportunities open to them, their calculation of profits and loss from the adoption of the technical packages promoted by the project, and their attitudes and feelings toward the implementing agency. The participant observer tries to become an insider without losing his or her status as an objective outsider.¹

There are several advantages to using participant observation in monitoring and evaluation. Obviously, the most important is that a phenomenon or process is observed in its most natural setting: the way that decisions are made in village cooperatives, the technical advice actually communicated to farmers by the extension worker, and the daily operations of credit institutions in rural settings. The resulting depth of insight is not easily obtained in any other way.

Moreover, participant observation helps reveal a phenomenon or process in its totality, as opposed to the partial reconstruction provided by interviews after the fact. For example, in an interview an extension worker

1. Strictly defined, a participant observer *participates* in the activities that are the subject of his study. Increasingly, however, the definition is used to embrace a lengthy residential observation with only incidental actual participation.

will describe the extension service from his vantage point; his description represents perceptions of his own motives, roles, activities, and experiences, or those of the farmers he deals with. But participant observers see a fuller picture; in addition to conducting interviews, they observe the actual behavior of extension agents, the reactions of client farmers when the extension agent has left, and the social and economic setting of the interactions.

Finally, participant observation reveals behavior patterns, social and economic processes, and environmental factors which the informants themselves may not be aware of or are unable to adequately describe. Participant observation thus is particularly useful in gaining insights about the conditions, needs, and behavior patterns of the rural poor and other vulnerable groups who are usually not able to articulate their problems and predicaments. An illiterate, old, widowed woman farmer in Lesotho does not find it easy to explain her problems and needs to an enumerator unknown to her, but a perceptive observer may see them clearly after spending a few days in the field.

The discussion of this method here is confined to those aspects which have special relevance to monitoring and evaluation. In this context, time is definitely limited; studies must be completed within weeks rather than months. Also, multisite studies may be needed to cover adequately the types of communities and organizations in a project area. More academically oriented studies are likely to involve longer residence at a single site.

Conceptual Framework and Data Requirements

The conventional wisdom in participant observation is that observers should not enter into the field with a preconceived conceptual framework because they tend to focus on the variables included in it and ignore other factors and conditions. Observers, it is said, should go to the field with an open mind and develop their frameworks purely on the basis of their field experience.

Such a posture is not justified in monitoring and evaluation studies for three reasons. First, there are the time constraints already mentioned. If fieldwork is conducted without even an embryonic framework, considerable time may be wasted in gathering an eclectic range of mostly irrelevant data. This is one of the reasons why project social analyses that rely on participant observation have often been disappointing.

Second, if several observers cover several sites, they must share an agreed framework. In its absence, they are likely to move their hypotheses in different directions and to generate information and explanations that are not comparable.

The third and probably most compelling reason is that monitoring and evaluation staff are not in the business of theory construction. Their ob-

jective is to point the way to the solution of implementation problems or to gain knowledge that can be used to plan future interventions. A conceptual framework is an aid and not a hindrance in this context.

The framework developed by the observers before entering the field should list principal issues and indicate the hypothesized relationships among them. In some cases, the main concepts should be defined so as to avoid later confusion. However, the conceptual framework for participant observation should not be as elaborate as that used in quantitative investigations.

Three general suggestions can be given. First, the number of variables should be limited; only those variables which are most pertinent should be included. The very purpose of a framework is defeated if an attempt is made to make it all-inclusive, and it is then of limited use as a guide in the field. Second, a careful review of the literature, and possibly qualitative interviews with key informants, should be undertaken before the framework is formulated. In a multisite study, the observers for different sites should be involved jointly in developing the framework. Third, a graphic representation of the framework facilitates the mapping of interrelations among the variables.² Example 6 provides an illustration.

EXAMPLE 6. Developing a Conceptual Framework for a Participant-Observation Study

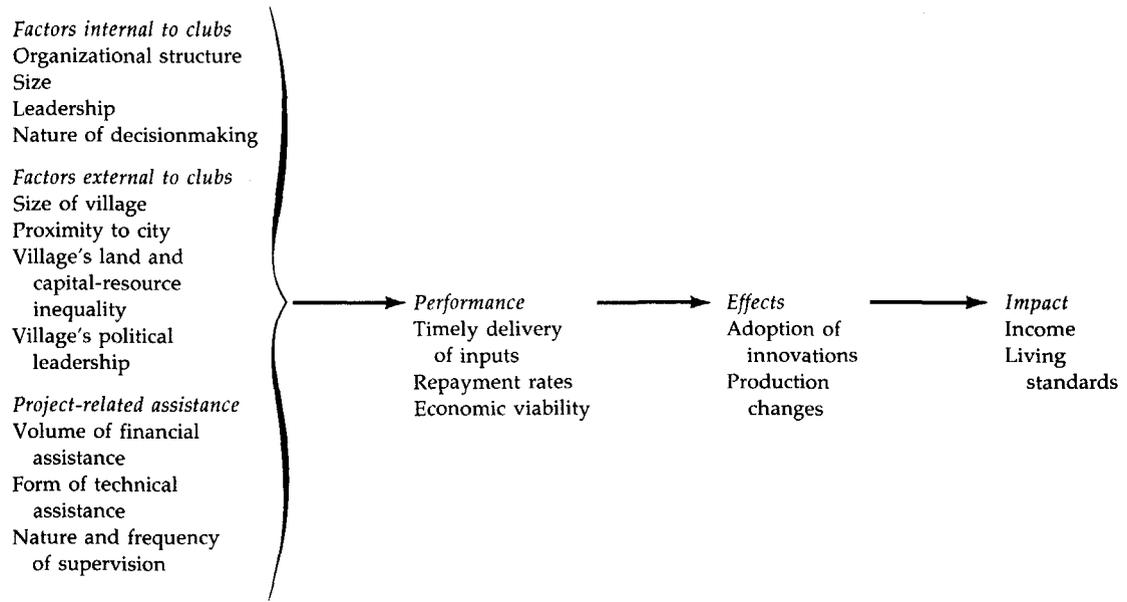
Farmers' clubs, organized as part of a project, secure short-term loans to purchase agricultural inputs which they distribute among their members. Dues are collected from members at the end of the harvesting season, and the credit bank is repaid. The project managers require the monitoring staff to do a study, using participant observation, to examine the overall performance of the clubs and to determine the factors and conditions that seem to affect them. The staff should follow three steps in this connection.

The first step is to construct a conceptual framework, as shown in figure 1, which hypothesizes a set of internal, external, and project-related factors which affect the performance of the clubs, which in turn influences the adoption of agricultural innovations by farmers, which leads to increased production and eventually to higher levels of living. This framework is only tentative, and the investigator will revise it as more information is accumulated and a better understanding is obtained.

The second step is to define the main concepts in the framework. Take, for example, the size of the club. Is the size to be determined by the number of subscribed members, the working capital of the club, or a combination of both? Consider leadership. Who should be regarded as leaders? Does the term refer

2. Matthew B. Miles and A. Michael Huberman, *Qualitative Data Analysis* (Beverly Hills, Calif.: Sage, 1984).

Figure 1. A Conceptual Framework for a Study of Farmers' Clubs



to the elected officials or to persons who wield considerable influence on decisionmaking within a club, even though they do not hold official positions? In most instances these issues cannot be fully resolved until the investigator goes into the field, but they need to be taken into account at the planning stage so that they are kept in mind when in the field.

The third step is to identify the sources of information on each variable in the framework. It is obvious from figure 1 that not all the required data will be obtained by direct observation. Other sources will include club documents and records; qualitative interviews to investigate the leadership factor; and project records to obtain the data on input supply, repayments, and economic viability.

Site Selection and Timing

The next stage in planning a participant-observation study is to decide where the study is to be carried out, whether it will be based on one site or several, and if there will be one or several observers. Although a single-site participant-observation study has limited usefulness, it may serve the required purpose. Sometimes one site can be treated as a typical case. Consider a situation—perhaps extreme—in which all ten marketing cooperatives initiated by a project are failing to provide the services for which they were established; here even a single case study can generate interesting information and insights. A single site study is also justified when the case in question is unique, for instance a total success or total failure. In the marketing-cooperative situation, suppose just one cooperative society is operating efficiently, a bit of hope in an otherwise depressing picture. Certainly close observation of this society is likely to shed light on the factors and conditions responsible for its success.

As a general rule, however, single site studies should be avoided. The major limitation of single site studies is obvious: the so-called typical or unique case may not be so. The inclusion of at least two sites enables some comparative analysis to be undertaken and gives an indication of variations between sites. For example, participant observation of both successful and unsuccessful marketing cooperatives is likely to provide a better understanding of the diverse sets of factors affecting their performance than that of either alone. The number of sites must be small, however, because participant observation is costly, time-consuming, and demands high skills; two or three sites will suffice for most projects.

Participant observation studies normally use informal sampling procedures to select the sites. The rationale for this is given in the discussion of evaluations based on case studies in chapter 8 of the companion volume. For such a study to succeed, several site selection factors which are independent of the desire for the sites to be “representative” must be taken into account:

- The phenomena to be observed must be occurring within the site and on a sufficient scale to facilitate the observation process.
- The community or organization must be willing to accept the participant observer.
- The participant observer must be able to enter into the normal activities of the community or organization. This may limit the number of possible sites because of language and custom considerations.

Timing is critical in participant observer studies, because by definition the events are to be observed as they occur. Wrong timing may ruin the study. This contrasts with interview surveys, for which there is some flexibility in timing because the respondent can be asked to recall the recent past as well as current events.

The importance of precise timing is accentuated in agriculture projects because actions are determined by the season. Rural credit organizations receive most loan applications during the planting season, when farmers wish to purchase inputs. Labor inputs vary widely through the crop growing season.

Farmers and farm organizations also follow daily routines that are associated with set times. For example, credit institutions may accept loan applications only during morning hours. Farmers in tropical climates may go to the fields early in the morning and return home at noon. Observation periods must be selected to reflect such rhythms in work.

Understanding Fieldwork

The need to establish sound initial contacts is obviously of particular importance when the stay of the observer is to be of limited duration. Extended field observations can afford to build slowly maturing relationships, but with shorter studies first impressions will be crucial—dress, manner, and conduct must be carefully considered. The observer should function in a way that people “must come to see and accept the evaluator as a person (more or less like themselves) rather than as professionals.”³

There are varying levels of participation implied by this type of study. One extreme is for the observers to become integrated completely into the community or organization acting, living, and behaving like its members. Such a course is normally not possible in a project context. The other extreme is for the participation of the observers to not go beyond their physical presence. The obvious limitation of this approach is that

3. Lawrence F. Salmen, *Listen to the People* (New York: Oxford University Press, 1987), p. 115.

the investigator remains primarily an outsider. The members of the community or organization may not reveal their innermost feelings and opinions; indeed, they may not function in a natural manner once they know that they are being observed.

The desirable middle course requires the observers to be more than passive observers but less than full-fledged members of the group. Although the observer does not act like a member of the community or organization, he participates in its formal and informal activities. For example, he might assist extension staff in organizing field demonstrations and meetings, attend staff meetings, and offer suggestions. The advantage of this course is that the observer is able to maintain his independence and yet become useful to the community or organization he is studying.

Practically all organizations and villages have cliques, subgroups, or factions; thus there is always a risk that the observer will arouse suspicion or antagonism among the members of a particular faction because he is perceived as being friendly to those of another. A persistent problem faced by many experienced observers is that they tend to be regarded as close to the village elites, so that small farmers, women, and other deprived groups are not candid in expressing their genuine concerns and views.

An observer has three choices about relationships with various factions and social groups. First, he can form a close association with the group considered best able to provide the required information. For example, if the observer is studying extension services, he closely interacts with extension staff, who become his primary source of information. The drawback of this approach is that his findings may be biased because of his reliance on a single group. Although he comes to understand the problems and perspectives of extension workers, he may remain unaware of the views, concerns, and difficulties of the farmers.

Second, the observer can spend time with different groups in the community or organization. In the above example, he first stays with the extension staff, then with the client farmers, and finally with the village leaders. The practical problem with this strategy is that the switch may not be easy if the observer has become identified with the first group to which he attached himself.

Third, he maintains an independent status from the beginning of the fieldwork. He meets freely with the members of the various groups, is friendly to all and intimate with none, and thus presents the image of an unbiased observer. This strategy makes it possible for the observer to observe a wide range of phenomena and interview members of different groups without arousing suspicion and misunderstandings. Example 7 illustrates some of the steps required for conducting an evaluation of a development project using participant observation.

EXAMPLE 7. Steps in Conducting Participant-Observer Evaluation

1. Become familiar with the background of the project. What are the project's objectives? Why and how was the site or the target population chosen? How did the executing agency and beneficiary population come together? Sources for this information are documents and interviews with key representatives of funding and executing agencies and with beneficiary leaders at the time of the project's initiation.

2. Learn the general characteristics of the population group benefiting from the project. Determine the history of the area and of the people—their places of origin, reasons for coming to the area or project site, and length of residence.

3. Choose a place of residence with care. Live in a fairly central location; in an area that is being upgraded rent a place somewhat better than average, one that offers the basic comforts and a separate entry. For a single person it is best to live in association with a family, yet clearly independent of it, so as to blend more easily into and become a part of the community.

4. Get to know the leading actors in the project well. The important contacts will be with the beneficiaries in general, their formal and informal leaders, and the main administrators of the implementing agency. Attempt to keep relations with each of these groups somewhat discrete. The manner of relating should be adapted to each—more professional with project administrators, less so with community leaders, and more informal with beneficiaries. Although everyone will know the reason for the observer's stay in the neighborhood, the official nature of that stay should be most apparent to the project personnel, with whom a certain degree of personal distance must be maintained to avoid bias, real or alleged, intentional or unconscious. The leaders will also be aware of the motive for the observer's presence, especially at the outset but less so over time. In some cases it may help credibility to have a letter signed by a government authority—not the executing agency but a more distant, neutral body. The people, however, should view the participant observer primarily as a neighbor and, to varying degrees, as a friend. At this level the relationship is far more personal than professional, although the people should be duly informed of why the observer is living among them.

Cultivate a few close contacts in diverse segments of the population representing various income groups, political factions, or other significant entities (such as youths, female heads of households, owners, and renters). Never be overly identified with any one group, but remain open and accessible to all: diplomacy at the neighborhood level.

Attempt to participate in community organizations and activities but without becoming overly committed. The observer's goal should be to demonstrate his involvement and have his efforts and interest appreciated, but to retain his independence.

5. *Inject issues of concern into discussions with residents after winning their confidence. These issues will have been identified with funding and executing institutions before the observer's entry into the community and will be refined by the participant observer in conjunction with the executing agency while at the project site. These issues should be introduced into conversation, or spontaneous discussion of them encouraged, and the talk should be guided to focus on the project—all in as unobtrusive a manner as possible. The aim is always to serve as catalyst and to provide the most honest response possible.*

Source: Lawrence F. Salmen, *Listen to the People* (New York: Oxford University Press, 1987), pp. 130–31.

Study Instruments

Written aids will be required in gathering and recording the results of observations and interviews so that important facts are not omitted or forgotten and so that the insights are noted and ordered to facilitate the preparation of reports. The observer may use the following.

- *Interview guides* are widely used in participant observation studies in a similar manner to that discussed in the two previous chapters.
- *Observation record forms* are similar in purpose to interview guides but are used to guide the eye and ear rather than the tongue; that is, they help the observer focus on events and existing conditions. One such form will be needed for each type of activity and situation included in the study. Example 8 shows a record form used in observing the proceedings of a cooperative society meeting. These forms usually should be filled in after the observation on the basis of notes taken at the time.
- *Document summary sheets* are prepared for reports and other documents that need to be condensed and recorded for reference at the stage of data analysis. A summary sheet should contain a brief description of the document, a summary of relevant points, and follow-up, if required; see example 9.

EXAMPLE 8. Observation Record Form for General Meetings of a Multipurpose Cooperative Society

<i>Seating arrangements</i>	<i>What were the seating arrangements in the meeting? How were the seats arranged? Were women sitting separately from men?</i>
<i>Attendance</i>	<i>What was the attendance? How many and what proportion of members attended the meeting?</i>

50 *Participant Observation*

<i>Conducting the Meeting</i>	<i>Who conducted the meeting? What format was used?</i>
<i>Agenda</i>	<i>What items were on the agenda? Were all of them covered in the meeting?</i>
<i>Discussions</i>	<i>Were there free and frank discussions? Which questions were discussed extensively?</i>
<i>Participation</i>	<i>Did the majority of those present participate in discussions? Did a few individuals dominate the meeting? If so, who? Did women participate in deliberations?</i>
<i>Interest</i>	<i>Was there keen interest in the proceedings? Were some members bored or indifferent?</i>
<i>Decisionmaking</i>	<i>Were the decisions unanimous? Were votes taken on some items? If so, which?</i>
<i>Factions</i>	<i>Did members appear to be divided into various cliques and factions?</i>
<i>Record keeping</i>	<i>Did anyone take minutes of the meeting?</i>
<i>Other items</i>	<i>Additional items which struck the observer as relevant.</i>

EXAMPLE 9. Document Summary Form

<i>Date received:</i>	<i>August 24, 1986</i>
<i>Source:</i>	<i>President of the multipurpose cooperative society in Arzung</i>
<i>Description:</i>	<i>Minutes of the meeting held on July 12, 1986</i>
<i>Summary of relevant information:</i>	<i>The minutes list the agenda and the specific decisions made in the meeting. Items 4 and 5 are quite revealing. Item 4 shows that there was opposition to the chairman of the society. The incumbent won by a narrow margin of two votes. It also indicates that the president of the society supported another candidate. Item 5 indicates that charges of financial mismanagement were levied against the manager of the marketing unit.</i>

Follow-up: *Should get more information about the cliques and divisions in the society. Should interview some members.*

Should find out if any action was taken to investigate the charges.

Should get copies of all the minutes for the last two years.

The notes taken with the aid of the above instruments will vary in form according to the observer's own style. But the record should cover each of the following three categories. First, the description of a phenomenon should be based on direct observation. The conventional wisdom in participant observation is that one should be as comprehensive as possible. Every detail should be noted, for an item which might at first appear to be trivial and insignificant may provide important clues to the understanding of phenomena or open new paths for further inquiry. Too many notes, however, will result in an accumulation of massive amounts of irrelevant data. With time constraints always in mind, it is important to focus on the important issues even if some interesting details are missed. Second, the investigator should record his own reactions, ideas, and insights that occur during the observation process. Third, any implications and tentative conclusions from the accumulated data should be noted as they occur to the observer. In participant observation, analysis and interpretation start during the field stage of the study. The observer's notes thus should examine the implications of his findings in terms of the hypothesized relationships. In many cases, revised hypotheses will be taking shape in the observer's mind as his observations continue, and this constant revision adds power to this type of study.

Minimizing Biases

The participant observer should be careful about two major sources of bias: the effects of the investigator on the observed situation and the effects of the observed situation on the investigator. The presence of an outside observer may disturb the situation in a community or organization. Leading people may become cautious in their remarks and behavior. For example, extension workers may become more patient and painstaking in their dealings with farmers, or the loan processing clerk may make an extra effort to help clients when an observer is present. In technical language, such changes are known as "reactive effects": they are compounded in project situations because the findings of a study can affect the fortunes of the people and organizations involved. Thus the project staff and actual (not necessarily intended) beneficiaries try to present a very positive picture of their activities.

Three steps should be taken to minimize reactive effects. First, the observers should stay in the field as long as possible so that they come to be perceived as a part of the setting and people become accustomed to their presence.

Second, some of the interviews with key informants should be conducted in informal settings, where people are usually more relaxed. The observer, for example, can interview respondents in cafes, houses, or over dinner, where they will be less self-conscious and hence more candid in their answers.

Third, the observer should constantly compare the accounts he is getting with the descriptions obtained from other sources. For instance, he can check with farmers to find out if officials are normally as conscientious in dealing with them as they were in the presence of the observer.

The observed situation also affects the observer, who may be unduly influenced by isolated cases or may begin to identify so strongly with the group that his objectivity is lost. To minimize these effects, the following four steps are recommended.

First, to provide a wider perspective on an activity, the observer should interview not only the persons who are central to the activity but also those who are peripheral to it. For example, when examining the working of an input supply agency, the observer should interview, in addition to current staff and beneficiaries, former employees, farmers who do not purchase inputs from the agency, and other well-informed persons not connected with it.

Second, the observer should not think in terms of specific personalities but at an abstract level. For example, he should consider what a manager of the credit union does, and not what Mr. X, whose commitment has impressed him, does as the manager.

Third, the observer should keep his main focus of study in mind and not be distracted by isolated events or interesting personalities.

Fourth, continuous reexamination and retrospection is most important. An observer should reflect on the changes in his views, judgments, and feelings over time. For this purpose, the observer should keep a record of his own impressions, and in a multisite study he should compare these impressions with those of other observers.

Limitations

Three limitations on the use of participant observation in the monitoring and evaluation of agriculture projects should be recognized. First, participant observation requires a high degree of skill and expertise on the part of the observers. It is a mistake to assume that persons without formal training and field experience can successfully engage in such studies.

Ideally, the observers should possess the following:

- Familiarity with the social, economic, and cultural milieu of the area
- Background in social science research methodology
- The capacity to learn from listening to the people, which means both patience and openness to the ideas and views of others
- Sufficient maturity to discern the relevance of what is being observed
- The capacity to relate to the people.

Some of these skills can be acquired from formal training. But others can come only through working with accomplished observers.

The second limitation is that properly conducted participant observation is very time-consuming. Typically, it takes an observer four to five weeks in the field before he is able to settle down and start systematic data collection. Another eight to twelve weeks may be necessary to obtain the full insight required and draw the appropriate conclusion.

The time is reduced if the staff who will be observing are stationed in the field already and are familiar with the local social and economic conditions, cultural landscape, and—above all—the people with whom they will interact. Thus they do not have to spend months to establish rapport and understand the prevailing conditions. There is, however, a negative side to this involvement. Because they are part of the project team, although the community accepts them, it may never regard them as true “insiders.”

Third, participant observation cannot be used to study highly heterogeneous populations. If, for instance, a project covers areas that have different cropping patterns, land tenure systems, and ethnic groups, it would be extremely difficult to identify a few sites which could represent the population.

These limitations need to be considered carefully before a participant-observation study is embarked on. When well done, it provides a valuable depth of insight; when badly done, it can seriously mislead.

5 | *Structured Surveys*

A STRUCTURED SURVEY, broadly speaking, is a method of interviewing people to collect information in which a formal questionnaire is used to structure interviews. There are two principal differences between this type of survey and qualitative interviews (which were discussed in detail in chapters 2 and 3). First, structured surveys are designed to generate quantitative data. When information is supplied in an open-ended format, it is converted into one of a limited set of coded options for the purpose of statistical analysis and presentation. Second, the coverage of the interview is decided upon and standardized before the survey begins; no changes can be made by the interviewer during the course of the interview. The interviews for such a survey also are commonly conducted with a sample of respondents selected according to randomization procedures, although this is not always the case (the role of probability sampling is discussed in chapter 6).

Structured surveys are almost indispensable in monitoring and evaluating agriculture and rural development projects. They are used to study a wide range of subjects ranging from the composition of the target population, to its reactions to project stimuli, to its more general attitudes toward and perceptions of changes in production activities, incomes, and standards of living.

The companion volume stresses the value of simple surveys of beneficiaries which record in quantitative terms the penetration of the project and the response of those reached. This chapter concentrates on how to design and conduct such surveys. The logical follow-up, namely problem-diagnostic surveys, will likely use the qualitative methods described earlier or a mixture of these and structured surveys. One remaining main type of structured survey, the socioeconomic baseline survey of the project population, is briefly discussed first.

Socioeconomic Surveys: Baseline and Follow-up

A socioeconomic baseline survey of the target population is a necessary source of information for the appraisal of a project. Many agriculture and

rural development projects have been designed without adequate data on the economic and social characteristics of their target populations. Project designers are forced therefore to make many assumptions and construct many hypotheses, which frequently prove to be false or erroneous. A baseline survey designed to yield reliable data on those economic and social characteristics that a project is expected to affect can therefore help improve the design of the project considerably. It can improve the odds for successful implementation of the project by providing managers with more detailed information on which to base their activities.

A socioeconomic baseline survey can also provide benchmark data on economic and social variables that can be used to measure the effects and impact of a project during and after its implementation. The survey therefore should ideally be undertaken *before* the implementation of the project and needs to be funded before the project starts.¹

A baseline survey will be similar in coverage and questionnaire construction to what is termed a general household survey: household members will be listed by demographic characteristics; the types of farming and nonfarming activities undertaken by household members may be recorded; sources of income from on-farm agricultural production may be enumerated in some detail; nonfarm labor earnings by each member, or alternatively expenditures, may be recorded; data on the social or ethnic status of the household may be collected; information on selected indicators of the quality of life—such as access to clean water, number of school-age children attending school, and months of self-provisioning with staples crops—may be gathered; and, if appropriate, some health and nutrition measures may be included. A baseline survey will, however, differ from most household surveys in one important aspect: it will be project oriented and so will focus largely on those characteristics of the target population that the project is expected to change in some manner.

Extreme caution is necessary in designing and conducting baseline surveys; they demand special skills and can be both time-consuming and expensive, especially when conducted in rural areas. Unless carefully constructed and limited in scope, they may fail to provide needed data on the main economic and social variables in time to be useful to the project. Experience from many countries indicates that critical bottlenecks in the processing and analysis of data often hinder the timely completion of baseline surveys. Such delays occur even when computer facilities for data processing are available. It is thus important to collect only the abso-

1. Whenever it is possible to include potential members of project monitoring and evaluation (M&E) units in the conduct of baseline surveys, the survey can also serve as a useful training exercise and help speed up the establishment of M&E units when the project is initiated.

lute minimum of data and to construct questionnaires that simplify the processing of data as much as possible.

In order to measure—both during and after the project—the changes in social and economic variables that the project is expected to bring about, the baseline survey needs to be easily replicable. For some variables, the survey may need to be repeated annually (for example, the section on crop production), at the midpoint of the project (for example, the section on farm practices), and at its completion (for example, the section on assets). The design of the baseline survey will thus have to take replicability as an important consideration in order not to impose undue strain on monitoring and evaluation units.²

In most of the literature on evaluation, baseline surveys refer to a major socioeconomic survey. But the term can be applied to the first round of a survey of any type, if it is conducted before project-induced change occurs. If the intended beneficiaries are to be monitored over time in terms of, say, adoption rates, the first round of interviews may well include supplementary questions on such topics as farm size, crops grown, and household size so that the rates may be monitored against the background of these explanatory variables. Such surveys are discussed in the remainder of this chapter.

Planning a Structured Survey

Stages in the planning of a structured survey include:

- a. Identification of the precise data expected from the survey
- b. Design of the survey in terms of the analytical method that will be used, for example, time-series or cross-sectional analysis, and the method to be used in selecting survey respondents
- c. Choice of the concepts and definitions to be used
- d. Definition of eligible respondents—which types of members of the population are to be included and which are to be omitted from the survey frame before sample selection
- e. Construction of questionnaire, including pretesting
- f. Selection of the sample
- g. Choice of interviewing method
- h. Establishment of data processing and analysis requirements
- i. Preparation of reporting formats.

2. For a more extensive treatment of how to design baseline surveys, see the U.N. household survey manuals, which consider in book-length treatments the topics needed to be covered. An example: U.N. Department of International Economic and Social Affairs, Statistical Office, *Handbook of Household Surveys* (New York: United Nations, 1984).

Items d, f, h, and i are treated at some length in succeeding chapters. We take up the other items here.

Data Requirements

The questions to be addressed in the survey must be clearly specified through discussions with the data users. For beneficiary contact monitoring, such questions may include: Are project services reaching the specified target groups? Does acceptance of project recommendations vary by type of farmer? Are adoption rates moving according to design expectations? Are production benefits beginning to appear? Too many surveys have been launched within a project and repeated throughout its life without any clear agreement between the managers and the monitoring staff on how the data emerging from the survey will be used.

Once there is agreement on data requirements, a careful review of other data sources within and outside the project should be made. We once accompanied a survey team to the field armed with a questionnaire on use of credit only to find at the first project office that detailed files on each credit recipient were maintained by project staff during regular visits to the farmers. Collation and analysis of existing data was the need, not a supplementary sample survey. A review of other surveys, even if they do not provide the required data, may help in making the methodological and design decisions necessary for the new survey.

Survey Design

One of the commonest needs in project monitoring is for a survey that is repeated seasonally to measure response and adoption rates. This is termed a *longitudinal survey*—one that provides a time series of the measured variables and rates. It is obviously essential for the definition of variables and mode of interview or measurement to remain constant over time, but there are two options for sample selection: draw a new sample of respondents for each round or maintain the same sample over time (sometimes termed a panel survey).

If the purpose is to estimate change over time, the optimal design is the panel survey because the error variance of the change from one season to the next is reduced by a factor of $1-R$, where R is the correlation between the behavior of the same farmer in two seasons, as measured over all farmers. If the purpose is to obtain the best possible estimate each season, any sample properly chosen is as good as any other chosen in the same way, so that mathematically it makes no difference whether the sample is maintained or replaced. But in practice both the monitoring staff and the manager will be concerned that if the sample is to be maintained over many seasons and if an unlucky random draw is obtained for the first

round, resulting in a wrong estimate, then the error will be perpetuated throughout the life of the project. With independent annual samples, one may argue that if one of them is unusual it will be evident because of the unusual result, and the chance of a serious bias being present in all values will be reduced.

There are also practical considerations. Respondents in a panel survey may become resistant to the recurrent demands on their time. This has not, however, been a serious difficulty in developing countries, where rural people, particularly project participants, tend to be more cooperative than those in industrial countries. Indeed, once the panel members' collaboration has been obtained, reinterviewing the same panel may be easier than seeking the approval of a new sample. Some surveys, however, such as objective yield measurements, may cause real inconvenience to the farmer, in which case imposing on his good will for too long may be asking too much. Another practical consideration is the effect of repeated interviews on the respondent's behavior—the contamination effect. In the context of project monitoring, such effects are often weaker than expected. Experiments show that people remember little of the detail of an interview. Certainly they do not readily change their behavior just because someone asks them a few questions; if they were that easily influenced, the extension worker's task would be much easier! Nevertheless, questions testing knowledge which are expressed in the form "Have you heard of *X*" cannot reasonably be asked a second time because "*X*" was mentioned by name on the first occasion.

No firm recommendation can be given in the light of the above, but a popular compromise is to partially rotate the panel from one survey round to the next. If a proportion p is repeated, the factor reducing the error variance of the estimate of change becomes $1-pR$. A typical value of R for project participants might be 0.5, so if half the sample is rotated each round the reduction in error variance will be 25 percent. Rotating a third or half of the sample in each successive round is a reasonable choice.

The main alternative to the longitudinal survey is the cross-sectional design, which involves data collection in only one survey round. Respondents are chosen from each of two or more groups, and the same information is obtained for each group. The analysis will then involve comparisons of the means or rates of the various groups.

A common example of a cross-sectional survey is one that compares the performance of the target group with that of a control group (a group that has not been influenced by the project) according to a set of agreed indicators. Selecting a valid control group may be rather difficult (discussed in some detail in chapter 8 of the companion volume).

Another type of single-round survey is one that compares the performance of various subsets of the target group according to some selected classifying variable. For example, we may wish to determine

whether the likelihood of adopting a technical innovation is affected by farm size. Adoption rates will be calculated for a set of samples drawn from each of an agreed set of size groups. The relationship between adoption rate and farm size can be tested by using analysis of variance or regression techniques. Of course, the sample may be drawn without advance knowledge of farm sizes. If the sample is of sufficient size, a reasonable distribution of farm sizes should result, and a regression analysis of the whole distribution can be undertaken without arbitrary stratification into fixed size groups. However, if the distribution of farm sizes is highly skewed, such a sample may result in too few selections of large farms to allow a sound determination of the relationship to be undertaken. Prior stratification, when feasible, is therefore an advantage. More is said on this in chapter 6.

Concepts and Definitions

Some of the concepts used in surveys are simple and do not require clarification. All of us know the meaning of age and sex. In most instances, however, even when we use simple concepts, there is a need to define them precisely in order to avoid ambiguity at the stage of data collection. Other concepts do not have precise meanings: rich and poor or large and small when applied to farmers can have different connotations for different people. One enumerator may regard any farm under 5 hectares as small, whereas another may regard one of 3 hectares as large. Income is a notoriously difficult concept to define with regard to small farmers. Differences in the valuation of home consumption and family labor and in the treatment of changes in herd composition of livestock make it difficult to obtain agreement on the meaning of income, let alone clearly communicate its definition to enumerators and, through them, to respondents.

There is an important difference in the degree to which technical terms are used if we use a structured survey with junior enumerators rather than the qualitative modes of information gathering described in the previous chapters. One of the virtues and requirements of the qualitative approach is that the interviewer, moderator, or participant-observer allows the respondents freedom of expression. Because the respondents therefore will use colloquial terms, any attempt to impose rigidly standard technical terms would defeat the purpose of the qualitative methods of data collection.

When using enumerators to conduct a dispersed-sample inquiry, however, it is clearly necessary to impose standard terms (with unambiguous definitions) and methods so that all enumerators are following a common set of instructions in which they have been trained. This is a much more important issue than many survey designers realize. For example, if the household is the reporting unit, it matters a great deal that a sensible and acceptable definition of household is adopted. This importance is ac-

centuated further when a holding is identified through its links with a household. The number of holdings covered and reported on will largely depend on the household definition. In an African country, the findings of a longitudinal survey that was repeated annually showed a major decline in household size in two successive rounds because in the second round, for some reason, enumerators treated many wives of male heads of family as having their own household if they lived in a hut separate from their husband. This change in definition by the enumerators, whether intended or not, in turn produced a large increase in the number of holdings and a reduction in their average size, because the plots cultivated by the wife were now treated as a holding separate from that of the usual head of family. Such terms as these, and others of particular relevance in the project context, are discussed below in order to remind the reader of the accepted definitions (if they exist) and to make suggestions regarding their adoption in monitoring and evaluation systems.

Household

Popular though it is as a unit of selection for contacting respondents in a survey, there is no agreed definition of a household. The simplest definition, adopted by many with U.N. sanction, is "a household is a group of people who live and eat together." Some international experts have gone further, or in a sense less far, and said that a household is whatever the community believes it to be. We suggest, however, that this will not do in the context of agricultural project monitoring, where most households are operating a single holding. We need to define unambiguously the holding via the selected household. We propose the following:

A household comprises a person or group of persons generally bound by ties of kinship who live together under a single roof or within a single compound and who share a community of life in that they are answerable to the same head and share a common source of food.³

Actually, unlike many national surveys, project surveys do not so commonly involve the selection of households. It is more likely that the farmer or project participant is selected directly and his farm and household identified accordingly.

Holder

Holder is the technical term for what we call in common parlance a farmer. This word is widely preceded by the word "small," that is,

3. Dennis J. Casley and Denis A. Lury, *Data Collection in Developing Countries*, 2d ed. (London: Oxford University Press, 1987).

"smallholder," to indicate a farmer with a small farm. The word smallholder, despite its popularity in development literature, has no definition at all, other than the one just stated. A small farmer in one country such as the pampas of Argentina would be a very large farmer in Malawi. Ten hectares may be a minimum to eke out a living in arid areas of north-eastern Brazil, but it puts one among the most productive in the Punjab of India.

"Holder," however, has an internationally agreed definition which states that the holder is the person who exercises control over the operations of a holding and is responsible for the utilization of available resources. This, of course, begs the question, for we now have to define "holding" (see below). Before we leave the holder, however, we can say that in a project context in addition to being the person we normally describe as the farmer, he will also be described as the project participant and intended beneficiary. He is the person who gets the project's credit, buys the project's recommended hybrid seed and fertilizer, and so forth.

Holding

The holding is defined as a unit of agricultural production comprising all the land used completely or partly for agricultural purposes and all livestock kept and operated, without regard to legal ownership. The land should be in use for agricultural purposes, or having once been in use, be lying fallow with the intention of being used again at a later date. Communal land, land never used for agricultural purposes, and natural forest are not part of an agricultural holding. Although communal land is not part of the holding, livestock owned by the holder but grazed on such land is.

Note particularly that legal ownership is not the issue. The cultivation of the land or management of the livestock with utilization of the product is the determining feature.

Parcel

The holding is made up of one or more parcels. The number of parcels is an indicator of the degree of the fragmentation of the holding, for a parcel is defined as a piece of land operated as part of a holding that is completely surrounded by land belonging to other people's lands or communal lands. If a holding has three parcels, the farmer has three pieces of land that he operates that are separated from each other, sometimes by considerable distances. The existence of multiple parcels is a bane to the surveyor, who can never be sure that the farmer has revealed all his parcels. Moreover, if the physical measurement of crops is involved, remote parcels can present real difficulties.

Plot

There is no term more beloved of surveyors that has so little meaning to the person who is cultivating it than "plot." The common approximation is the concept of a field, which in many countries is more or less unambiguously definable because it is a part of a holding that is surrounded by a hedge, fence, or some other boundary marking. But in other countries such boundary markings are not used, even though the word or its local translation may still be used. The trouble with the concept of a field is that different cultivation practices may exist within it, and because many agricultural inquiries are going to be crop-specific, the field, even if it is unambiguous in definition, may not serve as the unit that we wish to record. So the plot is defined as a piece of cultivated land containing a single crop or single homogeneous mixture of crops. There is no necessity for plots to be defined by identifiable boundaries; the dividing line between two growing crops is the boundary of the two individual plots. Adherence to this definition can produce remarkable complications. Take, for example, a field planted with sorghum in which millet has been mixed in one portion and cowpeas further added in a subportion so that we have, despite the absence of identifiable markings, three plots: (a) sorghum, millet, and cowpeas; (b) sorghum and millet; and (c) sorghum. The plots denote, respectively, the portions of the field that contain each particular composition of crops. If physical measurements are to be taken for precise comparative yield studies, the plot must be strictly defined. In interviews with farmers, adherence to the definition will be impossible; the farmer will use his concept of a plot, which is likely to approximate a field.

Mixed Cropping

No issue complicates crop measurement more than that caused by mixed cropping. In common parlance, this term covers all situations in which a single plot contains more than one crop at some point during the growing season. Many distinct forms of mixed cropping occur, such as associated crops, which consist of an annual crop growing under a perennial; relay cropping, in which crops are added in a time sequence within a plot with some overlapping time periods when particular combinations are in the field at the same time; and the simple form of mixed cropping, in which two crops are interplanted throughout a roughly equivalent season. The measurement problems are discussed in chapter 7.

Project Participant

Project participant is a term we have used frequently and almost interchangeably with project beneficiary, thus expressing the hope that any-

one who participates in a project may obtain some benefit from participation.

Because many monitoring surveys focus on this group, it requires a careful definition. This is not always a simple task, as the following example illustrates. A popular form of extension service, known as Training and Visit, uses a set of contact farmers as the focal point for regular visits from an extension agent. The neighbors of the contact farmer, however, are also meant to be informed of the date and location of the agent's visit and are free to attend. Furthermore, it is hoped that the contact farmer will disseminate the messages delivered to his neighbors during normal social exchanges. Clearly, the contact farmer is a project participant in the sense that he participates in a regular dialog with the extension agent, whose services are funded by the project. But suppose the contact farmer finds the messages delivered irrelevant and does not adopt them. Is he still a participant? What about the noncontact neighbor of the contact who does hear the message and does adopt it? Is he a participant? In both cases the answer is yes, although for the former we have a nonadopting participant. The neighbor of the contact who shows no sign of adopting recommended practices is a difficult case, for ideally we need to determine whether his nonadoption is caused by rejection of the service or a lack of knowledge of it. One could expand these possibilities to even greater levels of potential confusion. In practice, however, we would describe all farmers who have some access to the messages or services as participants. In other projects, there is less ambiguity. A credit project provides credit to a listed number of approved applicants; these are the project participants. In practical survey terms, we will of course need to classify the general participants by type and level of participation, for example, by amount of credit used.

Recall and Reference Periods

In most interview situations the respondent is expected to recall certain events that occurred during a specified period that has already passed. The *recall* period is the elapsed time between the event recalled and the time of the interview. To interview a farmer at one harvest period, but also to ask him about his harvest in the previous year, is of course to ask him to recall an event that is one year old.

The reference period is that between defined "opening" and "closing" times for the event that is being recalled. The reference period should be closed; that is, its starting and finishing points should be clear to the respondent—for example, "yesterday." To inquire about labor in the last season's planting period, conversely, is to use an open reference period, for the respondent has not been given a precise definition of the opening and closing dates of the period, which in any case is likely to vary from farmer to farmer.

One end of the reference period is certainly closed if it is the moment of the interview. The problem usually is to define the start of the period in a way that is logical for the respondent. Open-ended reference periods are a principal source of bias in obtaining quantitative responses to questions requiring memory recall.

The other thing to remember is that, except for the most important events in life, memory fades with time. This issue is often glossed over by those who wish to conduct "before" and "after" surveys on the basis of one interview when it is already long "after." Memories of the "before" state of affairs are likely to be poor, and this often results in understated events and quantities.

Questionnaire Construction

The questionnaire is the most important element of a structured survey; it must be standardized, well tested, and list in a systematic manner all the questions which are to be put to the respondent. Several well-established rules and regulations for constructing a questionnaire are mentioned briefly below.

Open-ended versus Closed Questions

Questions in an interview can be posed in an open-ended or closed manner. Open-ended questions allow the respondent to give answers using his own language and categories. Closed questions confront the respondent with a set of predetermined responses, and he has to identify one or more of those that are applicable.

The main advantage of open-ended questions, as discussed in chapter 2, is that the answers are spontaneous. The respondent is free to speak his mind. He can suggest new categories or ideas that might not have occurred to the designer of the questionnaire. The main drawback of open-ended questions in structured surveys is that they create large problems for quantitative analysis. It is extremely difficult to code them. The coder has to read carefully all the responses, develop new categories, and code each case accordingly. Even then, he is not sure if he has interpreted the answer correctly. There is often a great deal of variation in the answers. The responses might range from one sentence to a lengthy explanation. Moreover, the recording of the interviews is more susceptible to the framework and judgment of the enumerator, who is likely to stress the points which in his opinion are important. And because many enumerators are involved in a typical survey, the reliability of the data suffers. Some enumerators might probe more deeply than others. Finally, the open-ended format is very time-consuming.

Closed questions help to clarify for the respondent the type of response sought—the listing of alternatives clarifies the question itself.

Moreover, closed responses can easily be coded and analyzed, and because interviewing takes less time, more items can be covered during a reasonable interview time. But closed categories may restrict the response. Sometimes respondents are not willing to reveal their ignorance regarding a topic, and they will arbitrarily select one of the response categories; others who have a viewpoint that does not fit within the listed options may opt for one of the stated ones to shorten the dialog.

Simple, matter-of-fact questions and questions that can be answered quantitatively can be adequately covered by closed responses. This is also true when the behavioral choices are limited. For example, it is very likely that most farmers borrow from a limited number of sources—relatives, friends, cooperatives, moneylenders, banks, and so on. Therefore predetermined categories can serve the purpose. Open-ended questions may be needed, however, if we want to understand why farmers prefer a moneylender to the village cooperative bank. In order to propose a list of offered responses for a closed question for which the choices are not obvious, an open-ended question can be used in a pilot set of interviews to determine the types of responses obtained.

One type of questionnaire that is often used—the tabular format—needs to be mentioned before the sequence of questions is discussed. When a series of numbers on a specific topic is required from the respondent, the enumerator may be provided with a two-way table in which to insert the appropriate numbers. Example 10 shows such a table for a livestock survey. A variation of this is the format used for enumerating households in which one row of the table is used for each individual and the columns refer to specific details about the individual such as age, sex, relationship to head of household, and occupation.

EXAMPLE 10. A Two-Way Table for Livestock Numbers

<i>Breed</i>	<i>Calves</i>	<i>Heifers</i>	<i>Cows</i>	<i>Bulls</i>	<i>Steers</i>	<i>Oxen</i>
Indigenous						
Exotic						
Crossbreed						
Total						

Tabular layouts are convenient for both the questionnaire designer and the enumerator, but the considerable latitude that is allowed to the enumerator must be realized and accepted. With such a layout, neither the form nor sequence of questions is specified, and the style of interrogation is very much at the enumerator's discretion.

For most surveys of the beneficiary-contact monitoring type, the questionnaire format is likely to be based on questions in a series, with perhaps an occasional small tabular insert among the questions. One com-

mon example is used in monitoring the penetration of an extension service—with local variants it opens in the manner shown in the extract in example 11.

EXAMPLE 11. *Agricultural Research and Extension Project*

MONITORING SURVEY: QUESTIONNAIRE

Form 1

Date _____
 Zone _____
 Number _____
 Contact/Noncontact (1) Contact (2) Noncontact
 Sex (1) Male (2) Female

A. CONTACT FARMERS

1. How many times did the extension agent visit you in the last four weeks? (0) (1) (2) (3)
2. During the last visit, how many farmers participated? (0) 0 persons (1) 1-5
(2) 6-10 (3) 11-15 (4) 15+
3. Have you told other farmers about the visits? (1) yes (2) no
4. Would you like to be visited? (1) more frequently (2) less frequently (3) same
5. Do you wish to remain a contact farmer? (1) yes (2) no

B. NONCONTACT FARMERS

6. Are you aware that an extension agent comes regularly to the village? (1) yes (2) no
7. Do you know where the extension agent conducts demonstrations? (1) yes (2) no
8. Do you know on which day he conducts demonstrations? (1) yes (2) no If yes, indicate day: _____
9. Do contact farmers discuss the extension agent's recommendations with you? (1) on a regular basis
(2) rarely
(3) never

10. Have you or a member of your household attended a demonstration in the last four weeks? (1) yes (2) no
11. If no, why not? (1) not interested
 [End Interview] (2) too far
 (3) don't know place or time
 (4) not enough time
 (5) other (specify)_____
12. If yes, who? (1) respondent
 (2) spouse
 (3) both
 (4) other (specify)_____

Form 2

Date _____
 Zone _____
 Number _____
 Contact/Noncontact (1) Contact (2) Noncontact
 Sex (1) Male (2) Female
 Crop _____

13. Total area under crop (hectares) _____
14. For each of the recommendations below, ask the following questions.

<i>Extension agent recommendation</i>	<i>Were you aware of it prior to agent visit?</i>	<i>Had you applied it prior to agent visit?</i>	<i>Have you applied it this year?</i>
_____	(1) yes (2) no	(1) yes (2) no	(1) yes (2) no
_____	(1) yes (2) no	(1) yes (2) no	(1) yes (2) no
_____	(1) yes (2) no	(1) yes (2) no	(1) yes (2) no
_____	(1) yes (2) no	(1) yes (2) no	(1) yes (2) no
_____	(1) yes (2) no	(1) yes (2) no	(1) yes (2) no
_____	(1) yes (2) no	(1) yes (2) no	(1) yes (2) no

Source: World Bank. This form is adapted from one as taken from Josette Murphy and Tim Marchant, "Monitoring and Evaluation in Extension Agencies Using the Training and Visit Approach" (Washington, D.C.: World Bank, forthcoming).

Wording of the Questions

As indicated earlier, in surveys using many enumerators it is necessary to apply standard concepts and definitions. The enumerators should be well trained in these, but the more technical terms should be avoided as far as possible in the actual wording of questions to be put to the respondent. For example, the question "What constrains you from expanding your holding?" involves not only the use of a technical word—

“holding”—but a formal word—“constrains”—that may be unfamiliar to the respondent. The interviewer must already have established the scope of the holding that is under discussion; the question would be better phrased: “Why don’t you increase your cultivation [or livestock]?”

It is very easy to word a question in an ambiguous way. Words such as “here,” “new,” and “currently” are open to various interpretations. When using even such an apparently unambiguous word as “you” it is necessary to be clear as to whether the question refers specifically to the person addressed or to any member of the household.

A question such as “What type of house do you live in?” is not clear, because there can be different interpretations of the meaning of “type of house.” It can be interpreted with reference to the size, location, facilities provided, construction material used, and so forth. Indefinite words such as “often,” “frequently,” and “many” lack precise definitions and must be avoided. In the context of structured surveys (with limited follow-up probing allowed), questions asking “why . . .” rarely give useful results. It is better to formulate possible reasons and frame specific questions on each.

If a list of options is to be read to a respondent, it must be short (three or four items) because the respondent has to group the individual choices. The alternative is to seek a yes or no response to each of a list of options, in which case the list can be longer and more than one yes is permissible. When allowing multiple responses to a range of alternatives, enumerators must be carefully instructed whether to conclude with a probe such as “Any others?” If so, and an additional response is elicited, the probe should be repeated until the respondent says no. Time spent on obtaining the best, simplest, and clearest wording of questions will be well rewarded in terms of quality of response.

Leading Questions

A question phrased so that it creates an impression that a certain answer seems to be expected is known as a leading question. For example, “How do you feel about the working of the village cooperative society?” is not a leading question, but “Would you agree that the village cooperative society is helpful to you?” is. The second question is likely to evoke a more positive answer than the first.

The use of emotionally potent expressions also leads to certain answers. Consider alternative questions for a survey in a Latin American society. First, “The government is stressing that the agricultural land should be redistributed among actual cultivators. How do you feel about it?” Second, “Critics of the government are suggesting that all agricultural land should be distributed among actual cultivators. How do you feel about it?” It is highly probable that we will get different answers from the same respondents for these two questions. Neither is recommended.

Double-Barreled Questions

Inadvertently, two or more questions are sometimes included in what is presented to the respondent as one. A good illustration is, "Do you think that the government should provide credit to farmers at affordable rates and assist them in getting the improved variety of maize seed at a subsidized price?" Obviously, there are several questions explicitly stated or implied here that should be asked independently. The problem with such questions is that they confuse the respondent, who happens to agree with one part of the question but not another. In this instance, a respondent who does not favor the idea of government providing credit but wants seed at subsidized rates would not know how to answer. The tendency with convoluted or double-barreled questions is for the respondent to relate to the last part and answer that. One of the signals that indicates the likelihood of a convoluted question is the inclusion of the words "and" or "or." The questionnaire designer should be on the lookout for these words and reexamine his question when they appear.

Response Set

A response set occurs when a respondent answers a group of questions in the same way regardless of the question content. It can occur when several questions are presented together with the same response format. For example, a questionnaire contains questions about technical packages, input supply, and agricultural marketing. The respondent is supposed to answer yes or no to all the questions. In such a situation, a farmer who is well disposed toward the technical package is likely to say yes to all questions if the questionnaire begins with technical package statements with which the respondent agrees. A response set can be avoided either by varying the response categories for each question or by avoiding placing together all questions referring to the same topic.

Question Sequence

Opening questions should be pleasant, interesting, and easy to answer. They also should stimulate the interest of the respondents in the survey. Dull or sensitive questions should not be posed at the beginning of the interview.

The flow of questions should be logical and smooth. The respondent should be able to see immediately the relationship between the questions asked and the stated objective of the survey. If the interviewer has told the respondent that the survey is being carried out to assess the credit needs of the farmers in the project area, and then asks questions about his family, the respondent might become suspicious about the ultimate objective. Inserting explanations within a questionnaire can help to smooth a transition from one topic to another. For example, after asking several questions about farming practices, the enumerator can say, "So

far we have talked about your farming experience and practices. I would like now to ask a few questions about your family."

The location of sensitive questions in a questionnaire is an important issue over which experts differ. Some believe that they should be put at the end. The obvious advantage of this course is that even if they cause a breakdown in respondent cooperation, answers on other matters have already been secured. Some survey specialists argue, however, that at the end of an interview the respondent is usually tired and bored and may give superficial answers because he does not want to prolong the interview.

We suggest the following two rules of thumb. First, sensitive questions should be introduced only when rapport has been established between the enumerator and the respondent; so they cannot be located at the very beginning. Second, they should be put where they are most relevant in reference to other questions. For example, questions about the nonrepayment of loans from a credit society should be placed where other questions about credit are asked and not with the section on farm practices or family size.

Verbatim or Shorthand Questions

If it is vital that the questions be delivered to the respondent in a particular form using precise terminology, then the wording must be given in the questionnaire as it is intended to be delivered—a verbatim questionnaire. If it is thought to be necessary to restrict the interview format in this way, then it follows that the questionnaires must be translated into all the languages that will be used during the interviews in various locations; otherwise the purpose is defeated. Such a questionnaire must also incorporate all possible sequences of answers so that the enumerator has a precisely worded follow-up depending on the response to a previous question. An excerpt from a verbatim questionnaire is shown in example 12.

EXAMPLE 12. A Verbatim Survey

4-0. Attitudes

4-1. *Did you or any other member of your household use the land now covered by the woodlot before the woodlot was started?*

(1) *Yes* (2) *No*

↓ ↓
Go to 4-2 Go to 4-4

4-2. *Has the closure of the woodlot created any difficulties for your household?*

- (1) Yes (2) No
 ↓ ↓
 Go to 4-3 Go to 4-4

4-3. In what way(s)?

- (1) Further distance to travel for grazing / collection of grass
 (2) Now purchasing grass and fodder to make up deficiency
 (3) No grounds for cattle to stand
 (4) Only some sections of the village population now allowed to use woodlot
 (5) Other (specify) _____

4-4. Do you agree with the use of this land for the woodlot?

- (1) Yes (2) No
 ↓ ↓
 Go to 4-5 Go to 4-6

Source: Roger H. Slade and J. Gabriel Campbell, *An Operational Guide to the Monitoring and Evaluation of Social Forestry in India* (Rome: Food and Agriculture Organization of the United Nations, 1987). Used by permission.

We prefer, when possible, to allow the enumerator some flexibility in the question and answer sequence even within the framework of a structured interview. The questionnaire then acts as a memory aid for the enumerator (as well as the means of recording replies), and the questions can be summarized neatly in almost shorthand form. It is advisable to avoid mixing verbatim and summarized questions in one questionnaire, for the enumerator will then tend to deliver the shorthand questions as if they were expressed in full. If flexibility is allowed, there is a danger that some questions will be missed.

Provision for Recording

If the reply to a question is a number or quantity, a precisely defined space for recording the answer is required. If the survey is to be processed with a computer, it is recommended that boxed spaces be provided so that one digit is entered into each box. Item specification that needs to be coded can be similarly provided for, as in the example below.

Crop	Crop code	Number of units	
		Harvested	Sold
_____	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/>
_____	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/>	<input type="text"/> <input type="text"/> <input type="text"/>

To classify the category of a response, the general format is to list the agreed choices with an instruction to the enumerator to tick or circle a number according to the response received. Two versions are shown below.

How many times was the plot weeded?

<i>Tick:</i>	<i>Circle the number:</i>
Once _____	1. Once
Twice _____	2. Twice
More than twice _____	3. More than twice

When probing the respondent's attitudes to a project-related issue, we may ask the respondent to place his opinion within a scale ranging from very positive to very negative. The question is posed, followed by the supplementary:

Do you:	Strongly agree	_____
	Agree	_____
	Disagree	_____
	Strongly disagree	_____
	(No opinion)	_____

We recommend providing an even number of options, say, four or six, although an extra one may be needed to provide for those who genuinely have no opinion on the issue. If an odd number is used, there is a tendency for responses to bunch on the middle one of the series. This is less likely when the choices are delivered verbally but occurs when the questionnaire is completed by the respondent personally.

When more than one response may be ticked for an individual question, it may be possible—and if so, useful—to ask the respondent to rank his choices in order of importance. For example, a question on farmer's clubs might appear as follows, together with the instructions to the enumerator.

What are the most important activities that your farmer's club should undertake?

Supply credit	<input type="checkbox"/>
Supply fertilizers	<input type="checkbox"/>
Supply seeds	<input type="checkbox"/>
Arrange for combined marketing	<input type="checkbox"/>
Arrange discussions	<input type="checkbox"/>
Other (specify) _____	<input type="checkbox"/>
_____	<input type="checkbox"/>

(*Note:* If the respondent gives two or more responses, ask him to say which is the *most* important to him, the next most important, and so on, and enter 1, 2, 3, etcetera, in the appropriate boxes.)

In general, the questionnaire designer should strive for a printed questionnaire with a neat, uncluttered appearance. The enumerator often is

required to complete the entries in far from ideal conditions for writing. A neatly laid out, well-spaced questionnaire is less likely to cause recording errors. Such a questionnaire also facilitates the entry of the data onto a computer file. Errors at this stage can be numerous if the questionnaire does not have neat boxes and all answers recorded on the right-hand side of the page.

Pretesting a Questionnaire

A newly prepared questionnaire should be pretested on a few pilot respondents in order to identify weaknesses, ambiguities, and omissions before it is finalized for the survey itself. Unfortunately, this stage is often neglected. Many questionnaires that have been used in surveys would have received further badly needed attention even if they had been tested on an office colleague. The test should detect the following problems:

- *Wording of questions.* Was the wording of the questions clear to the respondents? Did all respondents derive the same meaning from the questions?
- *Construction of sentence.* Were the sentences appropriate? Were they too short or too long? Did they give unnecessary details which confused the recipients?
- *Question format.* Was the question format suitable? If there were open-ended questions, was there a great variation in responses which would make them difficult to code? Were the existing response categories for the closed questions adequate in range?
- *Difficult questions.* Were there questions which the respondents found it difficult to answer?
- *Same answer.* Were there questions for which all the respondents gave the same answers? (This indicates that the questions were nondiscriminating.)
- *Refusal rate.* Was there a tendency for respondents to refuse to answer particular questions? (Fortunately, in general refusal is not a serious problem in rural areas, but a particular question may be disturbing.)
- *Time requirement.* What was the approximate time needed to complete the questionnaire? Did the respondent seem to tire at the end?
- *Interviewer's convenience.* Did the enumerators find it difficult to administer some parts of the questionnaire? Were additional instructions needed? Were the probes appropriate?
- *Coding.* Were there problems in coding the data?
- *Usefulness of the data.* Was the questionnaire able to generate the type of information which was expected of it?

Pretesting the questionnaire is a vital stage in preparing for a survey. Survey designers neglect it at their peril.

Interviewing the Respondents

Questionnaires need not be administered individually in face-to-face interviews; they can be mailed or, in industrial nations, where the majority of the populace has a telephone, conducted by telephone. But such methods are out of the question in most rural societies, and individual interviews undoubtedly remain the most practical mode of data collection. Even when the respondents are literate, this method has several advantages over other methods. Experience shows that a personal interview maximizes the chance of obtaining the respondent's full cooperation. When an item is not understood, the interviewer can repeat it and probe for details if necessary. The main limitation of the interview method is that it is expensive. A relatively large proportion of the budget is spent in providing the salaries and the traveling and daily expenses of the enumerators.

The role of the enumerator in a structured survey is quite different from that of his counterpart in a qualitative survey. In the case of the latter, the interviewer exercises considerable independence and initiative about the nature, form, sequencing, and types of questions. He is not in any way restricted by the interview guidelines he uses; it provides at best helpful hints and not detailed instructions. Enumerators in a structured survey must adhere to the questionnaire and the guidelines provided.

The enumerator should cultivate a neutral attitude toward the subject and keep his own feelings, opinions, and judgments to himself. In no case should he engage in any verbal argument with the respondent. If the enumerator has reason to believe that the respondent is not telling the truth, he should make note of it in the questionnaire so that his supervisor can take any necessary follow-up action.

Rural people seldom refuse to be interviewed, but sometimes they are unavailable when the enumerator pays his visit. Whenever possible, the enumerator should inform potential respondents of the date and time of his arrival and of the interviews. The convenience of the respondent must be considered. During the peak of the agricultural season, when farmers are extremely busy in the fields, long interviews should be avoided as far as possible.

The enumerator should identify himself and provide a short introduction regarding the project and the purpose of the survey. Many projects, however, have permanent enumerators who reside in local areas and are known to the respondents. Obviously, in such instances little self-introduction is required.

Although the enumerator should explain the purpose of the survey, he need not be very specific about it. Lengthy descriptions often confuse rather than enlighten the respondents. Moreover, long introductions can generate biases in the interview by predisposing a respondent to answer in a particular way. Brief noncommittal statements such as "We are trying to find out about the working of the extension service" or "We want to know about the availability of agricultural credit" are usually enough.

The general principles of interviewing given in chapter 2 apply to structured surveys, except that the interviewer is not allowed great discretion in developing the interview.

The enumerator should avoid asking those questions on the questionnaire which it is clear from earlier answers do not apply. For example, "How old is your wife?" should not be asked of a respondent who has indicated that he is single, nor "Did you use fertilizers for the wheat crop during the last agricultural season?" of a respondent who has indicated that he does not grow wheat. Inapplicable questions can irritate the respondent and undermine the credibility of the enumerator in his eyes. If the questionnaire designer has done his job properly, the risk of the enumerator making this mistake is minimized because the questions will be in a logical sequence and appropriate skip instructions will be provided. Nevertheless, the enumerator should stay alert to the possibility of asking an irrelevant question.

6

*Sampling for Monitoring
and Evaluation*

THE PURPOSE OF SAMPLING is to economize on the resources that are needed to collect and analyze statistical data. Instead of using information from all members of the population, one collects it from only a part of the population; this part is taken as representative of the whole. Sampling theory allows one to calculate potential errors and maximize efficiency in selecting the respondents of the study. The use of samples can also improve the quality of data because of improved management of surveys made possible by using a limited number of respondents, and hence enumerators.

To limit the length of this chapter, the examples given are based on estimates of rates and percentages.¹ Readers of the companion volume will know that we regard such estimates as fundamental to project monitoring. For treatment of more complex examples dealing with continuous variables such as income or yields, the reader is referred to standard textbooks.

Several terms that refer to various aspects of sampling need to be defined. The population of concern is called the universe, and the part studied is called the sample. Inference from the sample to the universe is made in practice through the application of an estimation formula which gives the estimated universe figure in terms of the sample observations. A formula of this kind is called an estimator.

In this chapter, notation for a quantity in the universe is indicated by a capital letter and for a quantity in the sample by a lower-case letter. Estimators are distinguished with a circumflex. Means (averages) are denoted by a bar above the symbol. Thus \bar{Y} is the universe mean, \bar{y} is the sample mean, and the equation $\hat{Y} = \bar{y}$ states that the sample mean \bar{y} is used as an estimator of the universe mean \bar{Y} —in other words, that the

1. This chapter is based on a previously published monograph: Christopher Scott, *Sampling for Monitoring and Evaluation* (Baltimore, Md.: Johns Hopkins University Press, 1985). It includes revisions by the author.

mean for the universe will be estimated, in the simplest possible way, as equal to the mean for the sample. Many estimators are as simple as this, but sometimes more complex estimators are chosen, notably if one wishes to bring in information from outside the sample.

Sampling error represents the uncertainty attributable to sampling because estimates are being made from a sample rather than the universe. For a given sample design and estimator, sampling error is defined as a combination of two components: sampling bias and sampling variance. Imagine that the process of selecting the sample is repeated many times and yields many independent samples, all of the same size and design. For a given variable Y , any particular sample s will provide an estimate Y_s , and these estimates will vary from one sample to another. If the mean of the estimates Y_s over all the possible samples is equal to the universe value of Y , the estimate is said to be unbiased; any difference from the universe value is called the sampling bias. The variance of the set of estimates Y_s obtained from the set of all possible samples is called the sampling variance.

In practice, sample designs and estimation formulas are chosen in such a way that the sampling bias is set at zero. Sampling variance is minimized to the extent possible within a given set of cost constraints. There are, of course, errors which would arise even if the whole universe were studied—nonsampling errors. These are survey errors: the respondent gives an incorrect answer or misunderstands the question, the interviewer makes a mistake in recording the response, an error is made when the data are transcribed, and so on. Such errors may have a random component similar to sampling error; but more important, they are likely to introduce an unknown bias.

Sampling theory allows the sampling variance to be estimated from the sample data generated by the survey. Formulas can be found in textbooks, and there are many computer software packages which perform the required computations. It is also possible to predict sampling variance if one knows the way the sample is to be selected, the sample size, and the variance of the variable in the universe. This allows one to choose the most efficient sample design.

Samples are obtained by repeatedly selecting sampling units. Such units may be districts, villages, holdings, persons, fields, plots, subplots, or other things. The selection process requires a *complete list* of units; such a list is called a sampling frame. A map may be a sampling frame, insofar as it is equivalent to a list.

One method of selection is to assign numbers to the units and then use a random-number generator (or a table of random numbers which has been produced by a random-number generator) to identify sample members one by one until the required number of units is selected. This procedure is called random selection. For a more common method of sampling

from a list known as *systematic selection*, units are selected at fixed intervals throughout the list, starting from a randomly determined point.

These methods give equal probability of selection to every unit in the list. This *selection probability*, also called the *sampling fraction*, is simply the ratio:

$$f = \frac{\text{number selected}}{\text{number in list}} = \frac{\text{number in sample}}{\text{number in universe}}$$

Suppose that the list (the universe) consists of project participants and suppose that, after selecting a sample with probability f and interviewing its members, we find that the number of adopters of some technique in the sample is y . We can estimate the number of adopters among all the program participants as y/f . More formally, we select as the estimator, $\hat{Y} = y/f$. If the percentage of participants in the sample adopting, namely, the *adoption rate*, is r , we can estimate the adoption rate R among program participants as r . Thus $\hat{R} = r$. Both \hat{Y} and \hat{R} are examples of unbiased estimators.

Inference from the sample to the universe, estimation of sampling error, and optimization of sampling efficiency all depend on sampling theory. This mathematical theory is based on the assumption that every member of the sample has a known and nonzero probability of selection. Sampling which satisfies this requirement is called *probability sampling*, or sometimes *scientific sampling*.²

Probability versus Informal Sampling

Besides its defining property—known and nonzero selection probabilities for all units—probability sampling is characterized in practice by several other features:

- Clearly defined selection procedures
- The use of lists (or their equivalent) as sampling frames
- The applicability of sampling theory
- The possibility of estimating sampling error.

Informal sampling is simply the complement: any sampling procedure that does not give specified values to the selection probabilities. There are many such methods, including quota sampling and purposive sampling.³

2. The term random sampling is sometimes used in this sense, but such usage can lead to confusion because the same term can be used to characterize what we have called “random selection” (in contrast to “systematic selection”).

3. Terminology in this area has become fluid in recent years. Purposive sampling once had a highly specific meaning (the sample was adjusted to yield a specified value of the mean of some important parameter), but it now seems to be used as a catchall term for any method in which the selection of units is subject to conscious purpose.

Most of these methods abandon the features mentioned above, although some retain the first one. Controversy over probability as opposed to informal sampling is usually the result of misinformation.

In the companion volume and in the earlier chapters of this volume, we have described specific project information requirements; some of these can best be met by probability sampling and some by informal sampling methods. It is the specific information requirement that determines the method, *not* the scale or complexity of the survey. This has been misunderstood by many who advocate rapid, informal methods. It is true that large-scale population surveys generally make use of probability sampling schemes, but this does not mean that large samples and full population coverage are necessary requirements for such sampling. Consider the following example.

A project manager believes that a particular message is achieving an adoption rate of 60 percent among project participants. He wishes to test this belief by interviewing a simple random sample of participants. To fix a lower limit of error and a confidence limit, he asserts that he will be satisfied if he can demonstrate with 90 percent confidence that the adoption rate is not below 50 percent. If the adoption rate is indeed 60 percent, a sample of 39 is needed to yield the desired confidence. (Simple random sampling is assumed, with $P = 0.6$.)

This example demonstrates two things. First, probability sampling in no way implies large samples; the sample size may be large or small, depending on the objectives. Second, probability sampling does not have to cover the whole population. In the example, the sample represents only the project participants. Probability sampling requires that a universe be identified, but the survey designer is free to choose the universe before starting sampling, and it does not have to be all-inclusive.

When a very small sample is contemplated, it can be argued that blind, random selection may produce an obviously unrepresentative sample whereas purposive selection can avoid this risk. Imagine, for the sake of simplicity, that the zone to be studied consists of a long stretch of terrain and that the average rainfall is high at one end, very low at the other end, and intermediate in the middle; imagine further that each of these three areas contains one-third of the population. If random sampling were used, it could happen that the entire sample would be made up of units from the area of high rainfall; a purposive selection would avoid this. This is precisely the type of situation for which stratified sampling was devised. With the technique of stratification, the zone is divided before sampling into three strata according to the three levels of rainfall, and a separate sample is drawn from each stratum. In this way the objectives of purposive sampling can be achieved while the advantages of probability sampling are retained.

Yet another frequently encountered misunderstanding is the belief that formal sampling is impossible unless one possesses certain statistical

information about the population. It is sometimes supposed, for example, that one cannot use a sample to compare the performance of large farms with that of small farms unless one knows how many farms there are of each size; or that one cannot select a two-stage sample, first of villages and then of farmers, without knowing from the start the number of farmers in each village. These views are mistaken. It is entirely possible to select a probability sample without having prior statistical knowledge of the universe; the only requirement is that a sampling frame exist or that such a list can be made. For example, from a list of villages one could, with no knowledge of their sizes, select a systematic sample of one in ten, then list the holdings in each selected village and select a systematic sample of one in twenty. This would give a formal probability sample of holdings with a uniform probability of one in 200. It is true that lack of statistical knowledge of the universe would bring certain disadvantages, notably that the sample size would be unpredictable and the sampling error would be relatively high. But these disadvantages, though real, do not compromise the probabilistic nature of the sample.

As we have seen, many objections to probability sampling are based in large part on misunderstandings. There is, however, one unmistakable disadvantage of formal sampling: the need for a list or sampling frame. Even here there may be misunderstanding: one sometimes hears it argued that because no list exists, or because the existing lists are seriously incomplete, probability sampling is impossible. This argument overlooks the fact that the sampler can arrange to make a list. Indeed, a listing operation is a typical component of most developing-country surveys, simply to provide the missing sampling frame. It is often unnecessary to list the whole universe of study; the technique known as cluster sampling allows one to limit the listing of holdings, for example, to a sample of villages. Similarly, if there is no list of villages, one may select a sample of communes or districts and make a list of the villages within that sample only. Despite these possibilities for economy, it remains true that the construction of lists will require time and resources, and this is a serious drawback of probability sampling.

It is important at this point to distinguish between constraints on time and on resources. If no list is available and there is such an urgent need for data that not enough time is available to carry out a listing operation, the time constraint is overriding. There is then no alternative to some kind of informal sampling. If the problem is one of available funds, however, the solution is less clear. It may be that greater economy would be achieved by diverting money from the main enumeration to the listing operation than by abandoning the listing operation altogether. A study conducted by the World Fertility Survey (WFS) found that if listing was omitted and the census was used to determine the selection of areas and the number of households to be selected in each area, the error involved

would be equivalent to a drop of 20 percent in the size of the total sample.⁴ It follows that if the listing could be done within a budget of 20 percent of the originally planned field cost, it would be more economical to reduce the sample than to abolish listing. The WFS study is based on conservative assumptions, and a figure higher than 20 percent might be more realistic in most situations. A suggested rule of thumb might be: if no list is available and if the creation of a list is limited only by cost constraints, it would be worthwhile to sacrifice a quarter to a third of the planned sample size in order to release funds to carry out the listing.

The cost of listing often will exceed this limit, particularly if the option of cluster sampling is not available and the overall sampling fraction is very low (making the ratio of listing costs to survey costs very high). In such cases the argument for informal sampling becomes strong.

What exactly are the disadvantages of informal sampling? Inability to estimate the sampling error is perhaps the best-known drawback, but it is far from the most important. The most serious disadvantage is the high risk of biased selection. If no list of the eligible units (universe) is made in advance, experience shows that certain categories of household or holding are systematically less likely to be selected: those that are inaccessible or remote, those with members who are frequently absent (seasonal migrants, for example), those that have been newly created, those that comprise a single person, those with members who belong to an ethnic minority (who are often regarded as not belonging to the village), and those with members who are socially or politically prominent (the enumerator may feel intimidated). Such biases occur in listing and enumeration even with formal surveys, but they are greatly augmented when no firm rules are given for selection. To some degree they arise because the enumerator, given a free hand, selects the respondents that present the least amount of trouble. To avoid this, some informal samples have been controlled by firm rules which seek to tie down the enumerator's choice. For example, the following instruction could be given: "Start at the village center (or chief's house, or mosque, or whatever). Walk east and interview at the tenth house. Continue, turning left and then right alternately every time an opportunity arises, always interviewing at the tenth house. Continue thus until the desired (specified) quota is achieved." Such methods work best in small towns. They may remove the bias caused by the enumerator's preferences, but they will certainly do nothing to en-

4. C. Scott and T. Harpham, "Sample Design" in *The World Fertility Survey: An Assessment*, ed. J. Cleland and C. Scott (New York: Oxford University Press, 1987). This study was based on ten countries (urban and rural sectors were analyzed separately) drawn from all continents. The finding takes account of the increase in sampling variance only; in practice there is likely to be a bias as well if the sample is selected without proper listing.

sure that dwellings in remote areas or in outlying hamlets are given the same chances of selection as dwellings in the main village.

Although these risks of bias are serious in every informal sample, they are only risks, not certainties. There is always a good chance that an informal sample will give a reliable answer. Moreover, every sample is a gamble—even a probability sample. The difference is that a probability sample runs a known risk, which can be reduced as low as one may wish, assuming the necessary resources are allocated. With informal sampling the risk is much greater and at the same time difficult to assess.

In chapter 8 of the companion volume we summarized the situations in which purposive sampling is appropriate. We have just referred to one: the excessive cost of preparing a list from which to sample with known probabilities. Two other situations need to be mentioned here to round off this discussion.

Exploratory diagnostic studies often need to be mounted quickly with no practical possibility of designing a formal survey within the short time allotted. Such a study may well be undertaken with informal selection methods when the diagnostic team reaches the field. Out of the study may emerge a clearer definition of the problem, a hypothesis regarding its cause, and even a suggested solution. The power of such a study is much augmented if a more formal survey can then be undertaken with the use of probability sampling methods to confirm the diagnosis for a larger population. Such a strategy is ideal for combining the flexibility of an unstructured search for ideas with the need for scientific confirmation.

The other situation in which purposive sampling is appropriate is the one discussed at length in chapter 8 of the companion volume—the case study to disprove a hypothesis or to describe the behavior of an exceptional performer. We quote from that section:

Case studies based on nonrandomized selections . . . cannot be used for valid inferences about the incidence of a phenomenon in the population or the average value of a variable in the population. But they can be used to disprove a null hypothesis that a particular constraint does not exist or a particular activity is open to all members of the population.⁵

In other words, when we wish to demonstrate that an assumption inherent in the project strategy is wrong, it is sufficient to identify cases in which it fails to hold. When we wish to show that benefits do ensue when the interventions are adopted fully and all conditions are met, it is appropriate to seek out a sample of such full, enthusiastic adopters.

5. Dennis J. Casley and Krishna Kumar, *Project Monitoring and Evaluation in Agriculture* (Baltimore, Md.: Johns Hopkins University Press, 1987), p. 128.

In sum, purposive sampling has a role when we are describing a phenomenon rather than wishing to infer its incidence in the population.

Sample Size

The first question likely to be asked when a sample survey is contemplated is "How large should the sample be?" In this section we examine the logic of this choice. Later sections will discuss how the sampler can make a specific decision in varying circumstances.

It is obvious that the larger the sample the smaller the sampling error. What is less obvious is the fact that the sampling error depends on the actual size of the sample rather than on the sampling fraction.⁶ The sample size needed to estimate the birth rate (at a given level of precision) in Gambia would be the same as that needed to estimate the birth rate in India. Roughly speaking, the relationship is a square-root law: to halve the sampling error, we must quadruple the sample size.

Sampling error depends not only on the sample size but also on the sample design. A multistage design increases the sampling error; stratification reduces it. These and other effects are described in later sections.

Sampling error also depends on the estimator used. By bringing external information into the estimate, we can generally reduce the sampling variance.⁷ In our discussions, it is assumed that the estimator for any mean, percentage, or rate in the universe is simply the same mean, percentage, or rate found in the sample, whereas if we wish to use the sample to estimate a total, we scale up the sample total in proportion to the sampling fraction. For such estimators, the sampling bias is zero, and therefore the only sampling error is the random error component.

Finally, sampling error depends not only on the sample but also on the universe sampled. In particular, if there is wide variation in the universe, sampling error will be high for a given sample size and design; if all units in the universe were equal, there would be zero sampling error and a sample of a single unit would suffice to give perfect precision.

In designing a sample, the investigator starts with a requirement for a certain degree of sampling precision. In other fields of study, one can usually specify the amount of precision needed by laying down acceptable margins of error. For example, an engineer will say "I need this component to be accurate to within 0.1 millimeter," implying 100 percent confidence that this limit will not be exceeded. But statistical distribu-

6. Strictly speaking, this is true only for sampling with replacement. In the more common sampling without replacement, the sampling variance is multiplied by a further factor, $1 - f$, where f is the sampling fraction. This causes a significant reduction only if f is substantial; usually it is negligible.

7. For treatment of this topic, the reader is referred to any sampling textbook. See, for example, Leslie Kish, *Survey Sampling* (New York: Wiley, 1965).

tions are not of this kind; they do not normally show a sharp cutoff. There is always a chance, however small, that the acceptable limit will be exceeded. One can never demand 100 percent confidence; the investigator can, however, specify the degree of confidence required. The closer this is to 100 percent, the larger the sample will have to be.

Thus the required precision in sampling has to be specified with two parameters instead of one—the margin of error and the confidence level. A typical specification would be, “The estimate must be correct to within plus or minus 10 percent with 95 percent confidence,” meaning that an error greater than 10 percent would occur not more than 5 times in every 100 trials. Typically, a confidence level of 95 percent is regarded as a standard requirement. But as we have argued in the companion volume, in the context of information to facilitate decisionmaking by project managers it is often appropriate to set a much lower confidence level, such as 80 percent. Relaxing the required confidence level brings down the required sample size substantially, as we will see below.

Obviously, the choice of the allowable margin of error will depend on the objectives of the inquiry. But it will also depend on the level of nonsampling error. In any data collection operation, some important sources of error are quite separate from sampling error. Grouped together under the term *nonsampling error*, these include a wide variety of aberrations: listing errors and omissions, interview nonresponse, response and measurement errors, interview recording errors, errors of coding and data entry, and programming or data processing errors. In the context of small-scale monitoring and evaluation inquiries, the most important of these are almost certainly the errors of response and measurement. Response errors arise primarily through memory failure, through uncertainties about units and dates, and through misunderstanding of questions. Measurement errors, in particular those of area and yield measurement, arise from the ambiguities inherent in the measurement task and the complexity of the operations. In practice, response and measurement errors can be considerably greater than most people imagine. For example, studies of bias in yield measurements have shown errors of 10–20 percent to be typical (see chapter 7). Reinterview studies have repeatedly shown the presence of alarmingly high levels of response error even on the simplest of survey questions.

The relevance of all this to sample design lies in the obvious consideration that one gains little by reducing sampling error if nonsampling error dominates the picture. Total error is obtained by summing the squares of the sampling and nonsampling errors. For example, if the uncertainty attributable to nonsampling error is measured by a standard deviation $\sigma_n = 10$ percent and if the sampling error is $\sigma_s = 5$ percent, then the total error is $\sigma = \sqrt{\sigma_n^2 + \sigma_s^2} = 11.2$ percent. Doubling the sample would reduce σ_s from 5 percent to $5/\sqrt{2}$ percent. If σ_n stays fixed, the total error

falls from 11.2 percent to 10.6 percent, an almost negligible gain. But such a doubling of the sample might well reduce the quality of supervision, which would result in an increase in the nonsampling error that could easily swamp this small reduction.

Although nonsampling error is difficult to predict and has no simple relationship to sample design, it is clear from the above that small samples often may be preferable to large ones. In general, once the sample size rises into the hundreds it is likely that efforts toward more thorough training and supervision of enumerators will pay off better than equivalent expenditures to enlarge the sample.

Single-Stage Sampling Techniques

In this section we give a formula connecting the margin of error, the level of confidence, and the sample size for the simplest types of sample design. The same chain of reasoning can be used in reverse. We may start with a given sample size determined by the available resources and use the same formula to determine the attainable margin of error for a given confidence level.

Many surveys need separate estimates for a number of distinct domains of study, such as individual project zones. The calculation for determining sample size that we describe has to be carried out for each domain for which an estimate is required, and the total sample will be the sum of these domain samples.

Consider the simplest type of monitoring problem involving sampling. It is assumed that a complete list of the universe elements—holdings, project participants, plots, or whatever—is available and that the purpose of the inquiry is to estimate the value of a simple indicator, which is defined as the percentage of the listed units that have a specified characteristic. For example, we may have a list of project participants and wish to estimate the adoption rate for some recommended technique, or we may have a list of all holdings in the area and wish to estimate the percentage that used a certain fertilizer.

We begin by assuming random sampling and then consider refinements. This method involves numbering the list and using random numbers to select a sample. Many handheld statistical calculators have a random-number generator. If the list runs to a three-digit number, every unit must be given a three-digit identification number (for example, the fifth unit will be 005, the fifteenth will be 015, and so on). Three random digits are then run off, and these are read as one number; the number determines the unit selected. If the random number exceeds the highest number on the list, or if it selects a unit that has already been selected, this particular three-digit number is rejected entirely (not modified). The process is repeated until the desired number of units is reached.

The procedure for fixing the desired sample size has been outlined in an earlier section. The investigator must decide on two things: first, the margin of error; and second, the percentage confidence with which it is required to know that this margin will not be exceeded.

The margin of error D , which is expressed in absolute percentage points, represents the largest acceptable error in the estimate. The margin may be two-sided or one-sided. If it is two-sided, one generally adopts equal values on either side so that the margins are $\pm D$. For example, if $D = 10$, then an estimate such as 37 percent will imply a confidence range of 27–47 percent. It may be decided that only gross overestimation need be guarded against. In that case, D will be one-sided, and if again we assume $D = 10$, an estimate such as 37 percent will now imply a confidence range of 27 percent and up.

Next, one has to fix the percentage confidence level for asserting that the margin will not be exceeded. It is here that the distinction between one-sided and two-sided margins needs attention. A figure of 90 percent confidence for a two-sided margin $\pm D$ implies a 10 percent probability that the error will exceed one or the other margin. This 10 percent divides into a 5 percent probability of going below the lower margin $-D$ and a 5 percent probability of exceeding the upper margin $+D$. Thus if the user is interested in one margin only, say the lower one, the requirement of 95 percent confidence that this will not be breached corresponds to a 90 percent confidence that the same two-sided margin D will not be breached. Confidence limits expressed in percentage terms have to be converted into a quantity K based on the normal distribution (see table 2).

To determine the required sample size, we still need one more parameter, namely, the variance V^2 of the variable of interest among the population. For a percentage rate R , assuming simple random sampling, this is $R(100 - R)$. We can now obtain the required sample size from the well-known formula $n = K^2V^2/D^2$, which becomes, for a percentage rate R , $n = K^2R(100 - R)/D^2$. The only unknown quantity here is R , the actual rate we plan to estimate. The formula is insensitive to the exact value of R , however, and a rough guess will suffice. In fact, $R(100 - R)$ reaches a

Table 2. Conversion of Confidence Interval to Normal Deviate

<i>Confidence level (percent)</i>		
<i>Two-sided interval</i>	<i>One-sided interval</i>	<i>Normal deviate (K)</i>
75.0	87.5	1.15
80.0	90.0	1.28
85.0	92.5	1.44
90.0	95.0	1.64
95.0	97.5	1.96

maximum value of 2,500 when $R = 50$; if no better guess can be made, this value may be adopted as a safe upper limit.

For example, using a simple random sample, we wish to estimate the adoption rate of a certain practice among the listed holders. We also want to know with 90 percent confidence that the rate will not be overestimated by more than 10 percentage points. We believe that the rate may be about 40 percent. We have a one-sided, 90 percent confidence limit. The conversion table shows that this corresponds to $K = 1.28$. D , the acceptable margin of error, is 10. The guessed value of R is 40. Thus, $n = 1.28^2 \times 40(100 - 40)/10^2 = 39$. Even if our guessed value of R is way off, the worse value ($R = 50$) would give $n = 41$, which is negligibly larger. In practice one might adopt a sample size of 50 to allow a margin for nonresponse and other mishaps.

A method that is both simpler than random selection and in general more statistically efficient (that is, it has a reduced sampling error) is systematic selection. The reduction in sampling error depends on the extent to which the list is classified in such an order that the study variable (here, the adoption rate) shows a steady trend as one goes down the list. This might happen if, for example, the farmers who joined the program first are listed first, or if the holdings are listed in order of their distance from the program headquarters, or perhaps even if the holdings are listed in order of size. The extent of the reduction in sampling error cannot be estimated in advance, but it would be reasonable to allow a somewhat smaller sample, say 45 instead of 50 in the above example, if the ordering of the list is known to be related to a factor which is likely to affect the study variable. If the list is not so ordered, there will be no reduction in sampling error but no increase either; so nothing is lost, and there is still the advantage that the systematic selection is marginally easier to perform (and to check) than random.

To carry out a systematic selection, number all the units in the list as before. Let N be the number of units in the list and n the number desired in the sample. Compute N/n to the nearest whole number. This will be the selection interval I . Select a random number A between 1 and I . The first selection is the unit numbered A ; the second is that numbered $A + I$; the next $A + 2I$. Continue thus, selecting every I^{th} until the end of the list. Check that the number selected is close to the predicted n .

If the list, or sampling frame, is not naturally ordered in any useful way, one could of course reorder it before starting the sampling. If it is a long list, this could be a tedious process. An alternative is provided by stratification.

Stratification

In stratified sampling, one begins by dividing the sampling frame into subsets called strata. Sampling is then carried out separately in each strata-

tum. This procedure tends to reduce the sampling error to the extent that each stratum is homogeneous with regard to the variable of interest. We therefore seek to create strata in such a way that the study variable (the adoption rate in our example) varies less within each stratum than between the strata. For example, one might classify farmers into those near the program headquarters, those at a medium distance from it, and those far away from it. One expects that the adoption rates will be high, medium, and low, respectively, in these three strata. If so, the amount of variation within strata is reduced; if not, nothing is lost. Like systematic sampling, stratification reduces error by ensuring that the sample is well spread out among the relevant subclasses of the population being studied. Stratification has several advantages over systematic selection. Perhaps the most important advantage is that stratification makes it easy to bring in many different stratifying criteria simultaneously. For example, one might make a three-way cross-stratification as follows:

<i>Distance from project headquarters</i>	<i>Large farms</i>		<i>Small farms</i>	
	<i>Enrolled early</i>	<i>Enrolled late</i>	<i>Enrolled early</i>	<i>Enrolled late</i>
	Near	A	B	C
Medium	E	F	G	H
Far	I	J	K	L

Here twelve strata based on three criteria are cross-classified. In general when stratifying, more is gained by using many criteria than by using many levels of one criterion. For example, if the above scheme were modified so that farm size was divided into six categories instead of two and the "distance from project headquarters" variable were dropped, there would still be twelve strata, but it is unlikely that this scheme would be as effective as the original in reducing the sampling error.

It is important that at least one unit be selected in each stratum. With a multiple classification such as the one diagrammed, it may be that the number of units in some cells will be too few to meet this requirement. If so, there is no objection to combining some of the strata. For example, in the light of the numbers existing in the sampling frame, we might decide to combine strata *A* and *E* into a single stratum and similarly to combine strata *K* and *L*, thus reducing the total number of strata from twelve to ten. We are free to make any such groupings before sampling begins.

Sample selection in each stratum must be carried out independently. The survey designer is therefore at liberty to use different sampling fractions in different strata. In sampling for adoption rates, deliberate use of different sampling fractions is unlikely to be worth the trouble (unless the main objective is to compare the rates for different strata; see below). Because the process of rounding off will produce some variation, however, and there may be significant variations in response rates among the strata, it is important to consider how the separate stratum results are put

together in an overall estimate. The simplest procedure is to compute the adoption rate for each stratum from the sample for that stratum; multiply this by the number of units existing in the stratum—that is, the number in the sampling frame—to obtain the total estimated number of adopters; sum these over all strata; and divide by the universe total to get the overall adoption rate. In symbols, this procedure would be $\hat{R} = \sum N_h r_h / \sum N_h$, where N_h is the number of units existing in the frame in stratum h , and r_h is the adoption rate found in the sample in stratum h .

The reduction in sampling variance achieved by stratification (assuming that approximately equal sampling fractions are used in all strata) is in essence equal to the ratio of the between-strata variance to total variance. It is unlikely that this can be estimated in advance. In examples of the kind we have cited, it would be reasonable to expect a gain equivalent to 10–20 percent—that is, the planned sample size could be reduced by this amount.

With stratified sampling, we can choose in each stratum whether to use random or systematic selection. The arguments in favor of systematic selection still apply.

Before we leave the topic of stratification, three common misunderstandings need to be cleared up. First, strata are not the same as domains of study. Domains of study are parts of the universe for which separate estimates are required. For example, one may wish to provide separate data for large farms and small, or for the northern and southern parts of the project area. Such domains are determined by the survey objectives; strata are determined by the need to optimize sampling efficiency. Typically, strata will be smaller than geographical domains of study. Often they are chosen in such a way that they do not cut across the geographical domains, but this is not essential.

Second, common belief is that scientific sampling requires that stratification be used. Third, another belief is that where the purpose is to compare two groups, these groups have to be made into two strata. Neither of these beliefs is correct. Stratification is not an essential feature of probability sampling, and two groups in the sample can easily be compared without prior stratification. Stratification is simply a device for improving sampling efficiency—that is, for allowing the same sampling error to be obtained with a smaller total sample. It achieves this by providing control over the size of the sample in each stratum; without stratification these sizes would be left to chance.

The Exit Poll: An Implicit List

One method of sampling, although probabilistic and formal, does not require any prior listing of the members of the universe. In certain circumstances, units belonging to a specific universe arrange themselves in

what amounts to a single file—for example, vehicles passing a given point on a narrow road or people coming through a door. In such cases the units are arranged in an implicit list. It is clear that a systematic sample may be selected by choosing every n^{th} unit which goes by.

For a monitoring service, the most likely application of this principle is the exit poll conducted on persons visiting a service center. For example, one might interview every tenth farmer leaving the center in a given week, starting on the first day with a random number between 1 and 10. The resulting sample may be taken as representative of the center's customers for that week.

Though simple, the method is slightly less simple than it appears, and four precautions are needed. First, it is necessary to decide who exactly is to be counted as a member of the universe. If the farmer comes with a friend or with his son, are they also to be counted as clients of the center? The answer will depend on the objectives of the survey. A decision has to be made and communicated in advance to the enumerator. Second, care is needed on the question of time coverage. It is unlikely that an inquiry based on a single day will be representative. Those who come on market day, for example, will have characteristics different from those who come another day. Even the most informal study should cover no less than three days. Similarly, it should not be assumed that morning and afternoon customers are equivalent: both should be covered. Third, if there is more than one exit, all exits should be covered. If the center provides a number of services, people using different services may leave by different exits. Sometimes exit polls from a market are attempted: here, the problem of covering all exits is even more crucial—and is likely to prove even more difficult. Fourth, there is the problem of fixing an appropriate selection interval. Ideally, the interval should be such that at the most crowded period the number of people coming through during one interview is equal to the selection interval. That way no one will be kept waiting for an interview and the enumerator will be kept as busy as possible. In practice this cannot be achieved with any accuracy; a guess will have to be made, and someone is likely to be kept waiting. Organization of the work will be easier if some person other than the enumerator can take responsibility for counting those passing and for informing the selected respondents—and if necessary for asking them to wait. It is possible to use more than one selection interval—say $1/10$ on market day, when there are more customers, and $1/5$ on other days—but if this is done, the market-day interviews must be given double weight in the analysis to correct the bias. If selection intervals are varied, this must be done in a systematic, preplanned manner so that the actual interval is known in relation to every interview. A method that is not legitimate is to have the enumerator select the next person to come through after each interview is completed. This would bias the sample in favor of times when visitors are few.

Sampling for Comparison of Two Rates

Project monitoring often involves comparing two groups, such as participants and nonparticipants or farmers in each of two locations or two categories. We assume below that the purpose is to compare two such groups in terms of a specified indicator, which is a percentage frequency among the units in each group. We also assume simple random sampling in each group (which implies that a list, or perhaps two separate lists, of units is available). If systematic sampling is used, a small reduction in sampling error can be guessed. If stratification is used, the formulas given should be applied within each stratum.

Let R_1 and R_2 be the actual rates in the two groups. The universe variance in each group, as we stated earlier, is given by $V_i^2 = R_i(100 - R_i)$, where $i = 1$ or 2 . We also saw that this variance is insensitive to the exact value of R . To a good approximation, unless R_1 and R_2 differ widely or are close to 0 or 100, we can take V_1 and V_2 as equal. On this assumption, it is easily shown that, for a given total sample size, the sampling error for the difference $R_1 - R_2$ is minimized when the sample is allocated equally between the two groups. We can now obtain a formula linking the sample size to the confidence interval. Assuming the two groups are equal in population variance and are sampled independently with equal sample sizes, the relationship is $n = 4K^2R(100 - R)/D^2$, where n is the total sample size ($n_1 + n_2$); R is the guessed average rate for the two groups; D is the margin of error for the difference ($R_1 - R_2$); and K is the normal deviate corresponding to the confidence level, as discussed above. It will be seen that this is simply four times the value of n given by the previous formula. (To detect differences, one has to contend with sampling errors in both the rates, and these add up; this doubles the variance, which quadruples the n required.)

For example, it is desired to compare the adoption rates for a certain practice between project participants and nonparticipants with a sample sufficient to detect, with 90 percent confidence, a difference of 10 percentage points in favor of participants. It is believed that the rates may be about 30 percent. What should be the sample size? Because we are interested only in a difference in favor of the participants, we use, say, a one-sided, 90 percent confidence limit. Referring to the conversion figures in table 2, we find that this corresponds to $K = 1.28$. The acceptable margin of error D is 10. The guessed value of R is 30. Thus $n = 4 \times 1.28^2 \times 30(100 - 30)/10^2 = 138$. Allowing a small margin for mishaps, we should therefore interview about 75 participants and 75 nonparticipants.

As stated, for minimum sampling error the sample size should be approximately equal in the two groups. It is unlikely, however, that the two groups will be equal in the universe. It follows that the sampling fractions in the two groups will probably have to differ.

It may happen that the cost of data collection is much higher in one group than in the other. Calculation shows that, if the same assumptions as before are maintained, minimum sampling error for a given cost will be achieved when $n_1/n_2 = \sqrt{c_2/c_1}$, where c_1 and c_2 are the costs per unit selected in the two groups. Thus if, for example, unit costs of surveying are four times higher in group 2 than in group 1, instead of dividing the sample equally one should select twice as many units from group 1 as from group 2.

Sampling with unequal sampling fractions in the two groups will present no problem if the groups are listed separately. If they are listed together but with an indication against each unit showing which group it belongs to, an easy method of proceeding is to number all members of group 1 in blue, from 001 upward, and all members of group 2 in red, also from 001 upward, and to carry out two separate sampling operations. It may happen that one does not know to which group the units belong until the interview. In that case it will almost always be best to use a single fixed sampling fraction throughout, sorting out the two groups only at the analysis stage. The reason is that nearly all the data collection cost arises from the time spent getting to the household, identifying the respondent, and launching the interview; much less time is spent asking the questions. If one has to interview a household anyway to find out to which group it belongs, little is saved by not asking the remaining questions.

When the sampling inquiry has been completed and the data are in, the formula given above can be used in reverse to compute the percentage of confidence attaching to the difference found. Let d be the difference between the adoption rates observed in the two groups. Replace D by d , and R by r (r = average rate observed) and use the formula to compute K . Look up this value in a table of the normal curve to obtain the confidence level. (If the samples in the two groups were not equal in size, replace n by $4 n_1 n_2 / [n_1 + n_2]$.)

Two-Stage Sampling Techniques

In a typical two-stage sample design, a sample of area units is first selected. A sample of households or holdings or project participants within each area, often called a cluster sample, is then selected.

Provided more than one unit is selected in each cluster, the use of two stages tends to cluster the sample, reducing the amount of traveling for interviews. This is the first advantage of two-stage designs. Second, if no prior list of units exists, the two-stage design reduces the work of listing: one has only to list units in the selected clusters.

The calculation of the optimal number of units to be selected within each sample cluster depends on the calculation of the intraclass correla-

tion δ , which is the correlation between responses given by members of the same cluster. This causes a loss in sampling efficiency expressed by the formula $z = 1 + \delta(\bar{m} - 1)$, where z is the relative efficiency of a simple random sample compared with a clustered sample and \bar{m} is the mean number of units selected per cluster. For example, a value of $z=1.3$ would mean that the clustered sample would have to be 30 percent larger than the simple random sample to achieve the same sampling error.

Despite its sampling inefficiency, cluster sampling brings a cost advantage, and this must be set off against the increased sampling error. For this purpose a cost function is required. The following model is usually adopted.

We divide the cost into three components. First, C_0 is a fixed overhead not affected by the sample design. Second, there are costs which are proportional to the number of clusters selected: costs of transport between clusters, of mapping, of listing, and so on. These are added together and represented by C_1 per selected cluster. Third, there are the costs incurred within each cluster which are proportional to the number of units selected in the cluster (the cluster take). These are C_2 per sample unit selected.

If n is the number of clusters selected and \bar{m} is the mean cluster take, then the total cost C is given by $C=C_0+C_1n+C_2n\bar{m}$. Putting this together with the efficiency formula above, we can compute the optimum \bar{m} using differential calculus. We obtain $\bar{m}_{opt}^2=(C_1/C_2) \times (1-\delta)/\delta$.

Typical field experience suggests a value of about 7 or 8 for the ratio C_1/C_2 when a listing of holdings has to be carried out, or about 5 when the list is already available. The value of δ varies according to the variable considered. For an adoption rate in an agricultural program, values of about 0.1 to 0.3 seems to be typical. Accepting $\delta=0.2$, and $C_1/C_2=7.5$, this gives $\bar{m}_{opt} = \sqrt{7.5 \times 0.8/0.2} = 5$ or 6. At this level, the inefficiency factor z becomes 1.9. In the example cited earlier, the required sample of 50 will thus need to be raised to about 95, so with an average cluster take of 6 we will need 16 clusters in the sample.

Because the purpose of clustering is to save transportation and listing costs, the smaller the area adopted as a cluster the better, at least until we come down to the village unit. Within the village, clustering would save almost nothing on transport and would involve extra work (because there are likely to be no within-village addresses).

In the discussion of single-stage sampling, we showed how the needed sample size can be calculated when comparing two groups. The formula was introduced with the proviso that the samples in the two groups be selected independently. If the samples are selected using a two-stage design and if both use the same area sample, the assumption breaks down. The usual clustering effect is likely to produce a positive correlation between the two groups, which will lower the variance of the difference

and hence lower the needed sample size. There do not appear to be any available data on the size of this effect for agricultural variables. If one extrapolates, precariously, from available data on social and demographic variables, one would expect a savings of perhaps 20–30 percent in the sample size. This will offset to some degree the usual increase in sampling error attributable to the inefficiency of cluster sampling: the latter increases the needed n by the factor $1 + \delta(\bar{m} - 1)$, where \bar{m} is now the average cluster take for either one of the groups.

For example, returning to the example used in the single-stage design, what changes would be involved if one decided to use a two-stage sample, with both groups selected from within the same clusters? For this example let us assume that δ is estimated at 0.1 and that 5 members of each group (10 in all) are to be selected from each cluster.

The previously computed n of 138 now has to be adjusted by two factors. First, for the clustering effect, multiply n by the factor $z = 1 + (\bar{m} - 1)\delta$, where $\bar{m} = 5$ and $\delta = 0.1$, giving $z = 1.4$. Second, for the variance reduction effect between the groups within clusters, multiply n by the factor 0.75 (a very uncertain estimate). Altogether, we obtain for the new sample size: $138 \times 1.4 \times 0.75 = 145$ approximately, a figure which might be raised to, say, 160, for safety. This implies 16 clusters with an average cluster take of five from group 1 and five from group 2.

Sampling for Rare Events

One sometimes hears it said that when sampling for rare events one needs a larger sample to achieve the same precision. But the situation is a little more complex than this.

Returning to the basic formula for computing sample size of a proportion, namely $n = K^2R(100-R)/D^2$, suppose we aimed for a two-sided margin of error of $D = \pm 5$ percent. If $R = 50$ percent, this implies a range of 45–55 percent. Putting the values into the formula, we obtain $n = K^2(50)(100-50)/5^2 = 100K^2$. Now if the events being counted are rare, R will be small. As an example, suppose $R = 10$ percent. Keeping the same value of D , we have an interval of 5–15 percent, giving $n = K^2(10)(100-10)/5^2 = 36K^2$, a much smaller sample for the same precision. Here, however, the precision is measured in *absolute* terms: D is the number of percentage points. If we wish to maintain the same *relative* precision, then a “5 percent margin” would mean 5 percent of the rate of 10 percent, so that $D = 0.5$ percent. For this precision a much larger sample would be needed: $n = K^2(10)(100-10)/0.5^2 = 3,600K^2$.

In summary, compared with sampling for common events, when sampling to estimate the frequency of rare events one needs a smaller sample to maintain the same absolute precision, but a larger sample to maintain the same relative precision.

A common situation encountered by the monitoring and evaluation worker is that in which an event is expected to be rare or nonexistent and assurance is sought from a sample inquiry to confirm this. For example, there may be rumors of a crop disease, or there may be concern over a possible misuse by certain farmers of a new technique. Often the sample study has been set up for some other purpose, but by adding one or two simple questions a check can be made on the matter of special concern.

Let us suppose that the sample yields not a single case of the event in question. Obviously, as long as the study is limited to a sample we cannot be completely confident that the event is not occurring. Let the true prevalence rate be R percent. How can we relate the confidence level to the value of R and the sample size, given that no cases are found in the sample?

Assuming a simple random sample, the probability of obtaining not a single case among a sample of size n is $P_0 = (1 - R/100)^n$. The larger the value of R , the smaller this probability. If we compute the smallest R for which the probability is below 0.05, we can say with 95 percent confidence that R is not greater than that value. The calculation has been made for selected values of n in table 3. Also shown is the value obtained if just one case is found in the sample. This is based on the probability $P_0 + P_1$, where $P_1 = n(1 - R/100)^{n-1} \times R/100$. As an example of the use of the table, suppose we have a simple random sample of 100 cases and that not a single case is found of the event in question; then we can assert with 95 percent confidence that the prevalence rate of this event is not greater than 3 percent. If one such case is found in the sample, we can assert with 95 percent confidence that the true rate is not greater than 4.7 percent.

Table 3. Upper Limit for R, at 95 Percent Confidence

Sample size (n)	Assuming no cases found in sample	Assuming one case found in sample
40	7.2	11.3
60	4.9	7.7
80	3.7	5.8
100	3.0	4.7
150	2.0	3.1
200	1.5	2.3
250	1.2	1.9
300	1.0	1.6

7

Measurement of Crop Production and Yields

IMPORTANT TECHNICAL AND OPERATIONAL ISSUES are involved in designing and implementing surveys to measure crop areas, yields, and production.¹ These surveys can be time consuming and cumbersome, and the “objective measurement techniques” which they use—such as crop cutting—frequently have had disappointing results. This chapter considers alternatives that involve the farmer in estimating production levels.

The appropriate system for obtaining data on the production of a project will depend on the precise use to which production estimates are to be put by project managers. These can differ from those needed by a central statistical agency, which has the responsibility to obtain aggregate statistics at the national and regional levels. The primary statistical need of a project is usually to compare specific subgroups of direct interest to the project. The subgroups could be those who receive different inputs, groups of beneficiaries who respond differently to those inputs, or control groups outside the project with which the project participants may be compared. In addition, the information on production that may be most useful to a project manager relates to the holding rather than specific plots.

As the companion volume emphasizes, beneficiary contacts are essential to implementing a project. Respondents should see themselves as participants in the activities of a project with a common interest in its processes. This opens the possibility at the level of the project of undertaking complex and repeated interviews, close observation and measurement, and in-depth inquiries about the reactions and responses of beneficiaries. None of these may be possible at the level of a national survey. But because of limited resources and constraints on the capability of the

1. This chapter draws on a monograph previously published in this series on monitoring and evaluation: C. D. Poate and Dennis J. Casley, *Estimating Crop Production in Development Projects: Methods and Limitations* (Baltimore, Md.: Johns Hopkins University Press, 1985).

monitoring and evaluation staff, detailed inquiries should be restricted to a very modest scale, and the beneficiaries themselves should be seen as a source of information. This is the case whether or not, for example, objective methods based on direct measurement give more or less accurate results than subjective methods. In many project contexts, the farmer's impression of what is taking place or has occurred is important in determining his reactions. His subjective estimates therefore are of real, objective value.

As is detailed in chapter 8 of the companion volume, the resources and time available for a project are often insufficient to measure production trends and hence to assess the impact of the project with any degree of confidence. Moreover, it is extremely difficult to determine the causality of any change in a rigorous manner.

A related special feature of project-specific evaluation of production changes is the use which can be made of carefully conducted case studies in conjunction with the distributional information generated through the routine monitoring system. The latter may provide information on the relative sizes of subgroups in the population according to their adoption and application of project inputs and recommendations. The objective of the case studies, as detailed in chapter 8 of the companion volume, would be to measure as accurately as possible production changes for each of these groups. To the extent that the groups are internally homogeneous and the differences between them are relatively pronounced, an assessment of the impact of the project on production levels may be made on the basis of a small number of cases. And if the number of observations is kept small, it may be feasible with available resources to adopt elaborate methodologies for crop area, yield, and production measurement which involve close observation, physical measurement, and repeated visits to farming households. By contrast, much simpler methods will have to be employed when it is necessary to obtain data on production for large, more representative samples of project participants.

Area, Production, and Yield

Areas under particular crops, production during a specified period, and yield are three basic variables, each of which may be of interest in its own right. Since yield (Y) refers to production (P) per unit area (A) during a cropping cycle or agricultural year, they have this simple relation: $P = A \times Y$, or $Y = P/A$. Consequently, production can be estimated by measuring directly, and then multiplying, area and yield. Or total production and crop areas can be measured directly, with yield estimated as their ratio. Simple though this point is, it is often overlooked. Yields need not be estimated directly if production and area estimates can be obtained more easily. Moreover, if the total production of a holding is the impor-

tant indicator of project performance, and it can be obtained directly, individual area and yield measurements are not needed.

Many conditions can complicate the measurement of A , Y , and P when a crop is not grown solely in pure stand or is cultivated and harvested more than once during the agricultural year. For example:

- Planting of a crop may occur within a short, specified period, but its harvesting may be spread out over a long period.
- Production from tree crops which are irregularly spaced within a holding may have to be expressed in terms of production during a stated reference period per tree.
- Occasionally, a crop is incompletely harvested. A well-known example is cassava, which is sometimes grown as a reserve crop to be harvested and used only if needed in an emergency. Or a cash crop may be abandoned because of unfavorable market conditions.
- The basis for estimating crop production is the harvested area. This may differ from the area that is cleared or planted. Strictly, therefore, crop areas should be measured at the time of harvest rather than immediately after planting. But because this is often impractical, measurement at the time of planting is used instead.
- The yield (or production) to be measured needs to be based on an agreed definition. A distinction can be made between biological or gross yield, harvested yield, and economic yield. Biological yield is that obtained before any losses during and after harvest. Subtracting losses in harvesting from this gives the harvested yield. Deducting losses during such postharvest processes as cleaning, winnowing, and drying gives the economic yield.
- If the same or different crops are planted successively on the same land more than once during a given agricultural year, agreed procedures are required to present estimates of cultivated areas (and thus to estimate yields or production) because the total area cultivated includes some land which is used two or three times.
- Mixed cropping potentially can cause great confusion in measuring and reporting crop areas, yields, and production. There are three basic types of mixed cropping: (a) Two or more crops are planted within a plot with each crop planted at a lower density than would occur if it was planted alone. (b) One crop is added between the rows of another crop, which was planted at its normal density (also referred to as interplanted crops). (c) One seasonal crop is grown under a permanent crop (referred to as associated cropping). Clearly, the production of crops is a function not simply of plot area but also of plant densities in the mixture and the detrimental or beneficial interaction between the mixed crops. If areas for each crop are presented simply as an aggregate of the areas over which that particular crop appears, irrespective of mixture, the resulting

figures will appear misleading unless supported by other information. There are three alternative possible solutions in such situations:

- Allocate the total area to different crops according to criteria such as the relative density in that mixture, which ensures that the sum of areas allocated to individual crops is the same as the total cultivated area.
- Allocate to each crop an “imputed area” which it would have occupied in pure stand to give the same production or yield—as a result, the sum of areas allocated to crops will not add up to the total physical area cultivated.
- Concentrate on only one or a few principal crops, and for each give separate area and production figures for pure stand and a few common mixtures—for example: maize in pure stand; maize with other cereals, with beans and pulses, and with permanent crops; and maize in all other mixtures. This solution is selective but avoids the complexity and arbitrariness inherent in the first two.

Area Measurement

Estimates of areas allocated to particular crops are required for various purposes. For situations in which little prior information is available, approximate area estimates by cropping pattern can greatly facilitate the planning and implementation of project activities. More careful assessment, especially of areas newly brought under specified crops, may be required when expansion of the cultivated area itself is an objective of a project. Accurate measurement of area by crop is needed when yields have to be estimated from measured production, or when total production has to be estimated from measured yields.

The choice of an appropriate measurement technique will depend upon both the objectives of the project and various operational factors, such as configuration of land, shape of fields, types of crops and cropping pattern, available maps and materials, and available skills and resources. The most common techniques are briefly reviewed below.

Air or Ground Transects; Area, Grid, and Point Sampling

In most circumstances in developing countries, the holding is the basic unit of study for surveys of crop production carried out in the context of monitoring and evaluation. But for some purposes, such as a rapid assessment study of a new project area, an estimate of the overall spatial distribution of the crops without reference to individual holdings may suffice. Possibilities will depend upon the quality and scale of the available mapping material.

Air or ground transects, a simple method of rapid assessment, involve dividing the study area into a grid framework derived from maps or aerial photographs. The choice of grid size will depend upon the size of the study area, whether sampling or complete coverage is intended, the degree of homogeneity in land use patterns, the resources available, and the degree of accuracy necessary. The procedure is to locate a predetermined point in each square of the grid and then traverse a fixed distance in a random direction. Along this transect, observations concerning land use, cropping pattern, and so on are made at regular intervals, and the proportions of observation points falling under specified land use and cropping categories are calculated. The methodology can involve low-level aerial photography or ground enumerators walking the transects.

Grid sampling differs from transects in that after the study areas are divided into squares with an approximate grid size, a sample of squares is selected and areas by crop and land use are measured or estimated within it. This can provide more accurate information, but considerably increases the work load. A variant called area sampling involves dividing the study area into natural geographic segments delineated by identifiable natural borders (ridges, roads, waterways, and so on), rather than into simple squares of equal size. When these segments are numbered and any available ancillary information is recorded, we obtain a frame of area sampling units from which samples of segments to be investigated can be selected. Where such segments cut across fields, simple procedures can be used to associate whole fields with individual segments. Area sampling methods are commonly used in surveys in developed countries, but holdings generally provide much more suitable units for production surveys in the small farming communities of developing countries, especially for monitoring and evaluation purposes.

Point sampling involves selecting a number of points on a map or grid and recording the cultivation practice at each point. This method is open to large biases because great reliance is placed on enumerators locating and adhering to the precise selected points.

Area Measurement within Holdings

In some cases good records may be available on the size of holdings, fields, and plots from either administrative sources or past surveys. In others, farmers are able to provide sufficiently accurate information in either traditional or modern units. In many cases, however, it may be necessary to undertake physical measurement of crop areas despite the complex and time-consuming nature of the task.

Fields and plots often have irregular shapes and curvilinear boundaries. The first step in measuring an area is to transform it into an approx-

imate polygon with straight sides and to demarcate its vertices on the ground. The more irregular the shape of the polygon, the larger the number of its sides. Little is gained, however, by increasing the number of sides beyond, say, ten or so. The important thing when using a straight line to approximate a real curving or irregular line is to balance the little pieces of the plot left out by including other small areas which are not part of the plot.

Two methods generally are used to estimate the area of the polygon. The most popular and most accurate involves computing the area with a trigonometric or graphical method after first measuring the length and direction (bearing) of each side. The second procedure, called triangulation, involves dividing the polygon with n sides into $n - 2$ triangles, measuring the three sides (or two sides and the enclosed angle) of each triangle in order to calculate its area, and summing the areas of each triangle.

For either method, a certain degree of redundancy usually is introduced into the measurements to ensure accuracy. When taking bearings around the perimeter of the plot, two bearings are taken and averaged in the forward direction, say from A to B, and then backward, from B to A. All sides should be measured, even though theoretically the last side in a closed polygon is fixed by determination of the other $n - 1$ sides. Because of inaccuracies in measurement, the polygon as measured may not "close"; that is, after starting from a point A and measuring the length and bearings of all sides, we may not return to the same point but to a different point A'. Vector AA' is the closing error. If AA' is greater than a certain proportion of the perimeter of the polygon (say greater than 3 percent, which still allows for errors in the area estimate much greater than this), the entire measuring process should be repeated. Programmed pocket calculators are now commonly used to calculate the plot areas; the standard programs also reveal the closing error in percentage terms.² For the triangulation method, the whole process should be repeated by constructing an alternative set of triangles.

When only approximate data are required, enumerators may measure the boundaries of the plot by pacing. The average length of step should be calibrated because it may vary from one enumerator to another. A more accurate alternative is to use a measuring wheel, which is graduated and has a counter that registers the number of revolutions as it is pushed along the line to be measured. Measuring tapes are also popular.

The most commonly used instrument for measuring angles or bearings is the compass. Optical range finders such as the surveyor's dumpy level usually include their own means to measure angular traverses.

2. See, for example, Food and Agriculture Organization of the United Nations, *Estimation of Crop Areas and Yields in Agricultural Statistics*, FAO Economic and Social Development Paper 22 (Rome, 1982).

Yield and Production Measurement

There are various ways to measure crop production and yield. The most appropriate technique will depend upon the scale of measurement operation as well as on project objectives, conditions, and available resources. This section describes the following methods:

- Harvesting and measuring total output of plots, fields, or holdings
- Crop cutting over samples of subplots
- Sampling of harvest units following crop gathering
- Interviewing farmers to obtain estimates
- Making use of eye estimates by agriculture extension agents.

These methods are listed according to the ascending scale on which they can normally be applied. Those listed first generally are more appropriate for intensive measurement on a relatively small scale; those listed last are more suitable for extensive coverage on a relatively large scale which involves dispersed samples. Crop cutting has become a popular method of yield estimation and has been used in a variety of circumstances, yet its comparative accuracy and efficiency have not been carefully evaluated (the technical details of this method are given in a section below). Certainly the method is taxing and is widely recognized to result in overestimates of yield. An important objective of the following discussion is to provide some comparison of that method with other possible ones. A general recommendation which emerges from the discussion is that in small-scale, intensive diagnostic studies, complete harvesting, which avoids many biases and random errors associated with crop cutting, may be the best method. To estimate production with a high degree of sampling precision requires fairly large dispersed samples, which may be infeasible with any method that uses actual measurement. The most appropriate method in many such circumstances will be an interview survey of farmers; evidence is accumulating that under certain conditions estimates by farmers will not result in a larger total error (including sampling and nonsampling errors; see chapter 6) than that obtained using the crop cutting method.

Harvesting Total Output

Harvesting of a complete plot is not commonly attempted because of the volume of work. If there is a range of crops and a sample of even moderate size, weighing the total harvest is unviable. But under some circumstances—because of the variability and biases to which the various sampling methods are subject—total harvesting may be suitable and

may indeed be the only method which provides the level of precision required.

In a small-scale, intensive study of yield differentials by farming practice, good cooperation can be expected from project participants for three basic reasons:

- The participants are aware of the project and will not be unduly suspicious of a project-related inquiry.
- The field staff can assist the farmer in conducting the harvest—thus in effect providing free labor.
- Farmers are interested in obtaining a full weighing of their harvest.

Complete harvesting is clearly superior to the crop cutting method in statistical efficiency and accuracy of data for the following reasons.

- Crop cutting estimates are subject to much larger sampling variability. This is because of yield variability within plots. For instance, in studies in Nigeria and Niger, yield variation within plots accounted for 40–60 percent of the total variation observed, that is, of the same order of magnitude as between-plot variation:

<i>Country</i>	<i>Crop</i>	<i>Subplot size (square meters)</i>	<i>Within-plot Variation (percent)</i>
Nigeria	Sorghum	50	45
	Yams	50	58
Niger	Millet, 1982	30	40
	Millet, 1983	30	48

If field-specific estimates are required it will be necessary, therefore, to have several crop cuts per plot.

- It is well established that the crop cutting method is subject to overestimation biases which increase with decreasing subplot size.
- Crop cutting measures the biological yield but takes no account of harvest and postharvest losses. What is often more relevant in the project context is the economic yield as given by postharvest weighing which allows for losses that have occurred before.

Regarding the time required, although the work entailed in the crop cutting method is likely to be substantially smaller than in helping the farmer to harvest the whole plot and weigh the produce, the latter method is not necessarily impractical if the whole plots average less than 0.5 hectares. In a field study in Nigeria, for example, one enumerator during a given time was able to execute eighty subplot crop cuts or to handle complete harvesting in thirty-two plots. The work rates in the two cases are of the same order of magnitude if two to three subplot crop cuts are involved per plot.

Crop Cutting

The crop cutting method involves (a) identifying a plot, (b) demarcating within it at random a specified number of subplots of predetermined size and shape, (c) harvesting the crop within the subplots in accordance with specified procedures, (d) drying the crop to standard conditions, (e) weighing the output, and (f) computing the yield by dividing the product obtained by subplot area. The accuracy of the method depends upon a number of conditions. In the case of uniform dense cultivation the method would be more accurate than under conditions of extensive cultivation with large intraplot variation. Errors (generally overestimation biases) are introduced because of difficulties in deciding exactly which plants near the edge of the subplot are to be included or excluded from the crop cut. This edge effect will depend upon the ratio of the perimeter to the area of the subplot, that is, on its shape and size. Larger subplots have a smaller relative edge effect; and for a given area, circles and squares have a smaller edge effect than triangles. The following table gives a portion of perhaps the most widely quoted figures in this context (from experiments in India³). It shows overestimates with crop cutting for triangular subplots of different size.

<i>Size of crop cut subplot (square meters)</i>	<i>Percentage overestimation in yield</i>	
	<i>Wheat irrigated</i>	<i>Wheat unirrigated</i>
43.80	0.0	0.0
10.95	4.8	11.0
2.74	15.7	23.4

The bias declines with increasing subplot size; for a given size, it is smaller for more uniformly and densely cultivated irrigated wheat crops and larger for nonirrigated crops. In either case, the bias is portrayed to be negligible for subplot sizes of about 40–50 square meters. But this positive conclusion cannot be automatically generalized to other crops grown under different conditions.

Another source of bias which tends to increase with increasing subplot size arises because subplots must be selected so that they lie wholly within the plot. This distorts the probabilities of inclusion of areas near the plot border in the subplots. Insofar as the yield near the border is different from yields elsewhere, this introduces an error called the border error.

Perhaps the most serious bias in crop cutting procedures arises from the almost unavoidable tendency among enumerators to consciously or unconsciously avoid empty or patchy areas within plots while locating

3. P. V. Sukhatme, *Sampling Theory of Surveys with Applications* (Ames: Iowa State University Press, 1954).

subplots. If random coordinates result in choosing a subplot which is clearly inferior to surrounding parts of the plot, a conscientious enumerator feels it is unrepresentative and may alter the location. Unfortunately they do not feel the same inclination when a particularly lush part of the plot is selected.

Finally, as already mentioned, the method does not take into consideration harvest and postharvest losses.

Conclusive empirical evidence of the magnitude of these biases and their determining factors is hard to come by. But several studies are currently in hand because of the Nigeria results contained in the Poate and Casley monograph (cited in footnote 1 of this chapter). These results showed that the use of 60-square-meter triangular subplots resulted in biases of the order of 14 percent under well-supervised conditions—biases that were of the same order of magnitude as farmers' estimates made immediately before the harvest. Two recent examples, one from Africa (Zimbabwe, example 13) and the other from Asia (Bangladesh, example 14) give further indications of the problems involved.

EXAMPLE 13. Recent Study of Maize Yields in Zimbabwe

Crop farming in Zimbabwe can be divided into four distinct sectors: large-scale commercial (LSC), small-scale commercial, communal areas, and the newly created resettlement sector. The first and third sectors predominate. The LSC sector includes some of the most advanced and productive crop farming in the world. For more than twenty-five years, the Central Statistical Office of Zimbabwe has been conducting full postal censuses of all farms in this sector, and a reliable series of production statistics is available.

By contrast, the communal areas are characterized by small-scale farming. There are 900,000 holdings with an average size of approximately 2 hectares. Three-quarters of this land lies in regions with low and erratic rainfall. Because of technological, ecological, and management variations, therefore, maize yields in the communal areas are expected to be lower than in the LSC sector. Maize yields in the LSC sector vary from 3.7 tons per hectare in farms of more than 100 hectares to 2.2 tons per hectare in farms under 10 hectares. In the communal areas, yields can be expected to be even lower, perhaps no more than 2.0 tons per hectare.

No reliable series of production and yield data is available for the communal sector. Recently, estimates from a national sample survey have been obtained using these methods:

- *Crop cutting from randomly selected portions of two randomly selected rows*
- *Reporting by farmers in an interview survey*
- *Eye estimates by Agritex extension officers.*

Table 4. Maize Yields in the Small-Scale Communal Sector: Comparison of Various Methods and with the Large-Scale Commercial Sector, Zimbabwe, 1984-85

<i>Area</i>	<i>Large-scale commercial sector</i>	<i>Crop cutting</i>	<i>Farmers' reporting</i>	<i>Agritex assessment</i>
Manicaland	2.4	3.4	1.8	1.9
Mashonaland Central	4.7	5.1	2.6	2.9
Mashonaland East	2.9	4.4	2.3	2.2
Mashonaland West	3.3	3.5	1.9	2.1
Midlands	1.6	3.3	2.5	2.7
Masvingo	1.6	3.4	1.7	1.8
<i>Unweighted average</i>	2.8	3.9	2.1	2.3

Table 5. Maize Yields in the Small-Scale Communal Sector: Comparison of Two Methods, Zimbabwe, 1985-86

Area	Crop cutting	Farmer estimate
Manicaland	1.6	0.9
Mashonaland Central	4.0	2.8
Mashonaland East	2.8	2.1
Mashonaland West	3.0	2.2
Midlands	3.2	2.1
Masvingo	1.3	1.0
<i>Unweighted average</i>	2.6	1.9

Results for six regions are shown in table 4. For comparison, the large-scale commercial sector is also shown. Though full experimental verification based on complete harvesting and measurement is not currently available, it is quite clear that yields obtained from crop cutting are highly biased in the direction of overestimation. Those from farmers' interviews and extension agent reports are close to each other (within 10 percent in all regions) and are generally more plausible. A substantial margin of error or bias in these estimates themselves cannot of course be ruled out on the basis of the available information.

Because of the unlikely results given by the crop cutting method as shown in table 4, the comparative study was repeated in 1985-86 with much tighter supervision of the enumerators carrying out the harvesting. Once again estimates by farmers agreed closely with those of the extension agents; the comparison with crop cutting gave the results shown in table 5. The crop cutting method seems to have been implemented with much more care, but the results are still typically more than one-third higher than the farmer's estimate.

Source: Central Statistical Office, Government of Zimbabwe.

EXAMPLE 14. Bangladesh Postharvest Project

The following is an illustration from a carefully conducted study to estimate postharvest losses. Physical losses in the operation up to threshing are the difference between the potential yield of the standing crop and the actual yield obtained by the farmer after winnowing the threshed crop. The latter can be obtained directly by weighing the farmer's produce. The former can be estimated in various ways—standard crop cutting methods on subplots of different shapes and sizes, or more laboriously, measuring yields per hill from a sample of hills in the plot.

The work was carried out under close supervision by staff holding master's degrees. Procedures were introduced to ensure standardization of moisture content each time the grain was weighed, and of winnowing and cleaning methods. It is believed that research procedures were controlled more closely than would be possible under normal monitoring conditions.

Table 6 shows that estimates derived from crop cuts exceeded actual yields by approximately 20 percent. The overestimates from the hill-count method were somewhat lower but more variable (6–19 percent). In either case, they were too large to provide meaningful estimates of postharvest losses.

In the same context, comparisons were made between yield estimation obtained from farmers' interviews and crop cuts. Farmers' estimates tended to be lower; in view of the results discussed above, this could indicate that the farmers' estimates were closer to the actual yield.

The study included provisions for the direct measurement of losses. This involved carefully gleaning and collecting scattered grain during field stacking and transportation, rethreshing a portion of the farmers' straw, and so on. This work yielded average loss figures for cutting (1.5 percent), field stacking and transportation (each 0.5 percent), and threshing and winnowing (1.8 percent), that is, a total of under 5 percent. The conclusion is that, apart from this difference between biological and economic yield, crop cutting methods overestimated the yield by 15 percent or so because of other sources of error (bias), and this in the best possible working conditions. The relatively large coefficients of variation, especially with the hill-count method, point to another difficulty associated with large intraplot variability.

Source: Personal communication from Martin Greeley, Institute of Development Studies, University of Sussex, England, 1986.

Sampling of Harvest Units

A method of measuring crop production that requires neither a total harvest weighing nor a crop cut involves a sampling of farmers' harvest after it has been gathered, at the point where it is transported for storage or disposal. At the time of harvest the farmer is left to collect the produce in his normal harvest units such as sacks, baskets, bowls, bundles, or individual roots and tubers. The enumerator visits the holding and inspects these units, obtains an estimate of the total number of these units harvested and of the estimated harvest area, and selects a sample of the units to measure the average unit weight. The total output is obtained by multiplying the estimated total number of units by average unit weight. Yield is obtained by dividing the total estimated production by the area harvested under the crop.

Where applicable, this approach can be used on larger samples than is possible with the more taxing procedures described above. For the method to work, however, certain basic conditions have to be met: (a) the crop must be gathered in identifiable and complete units; (b) the units should not be too variable, so that the average unit weight can be estimated with necessary precision with a reasonable sample size; (c) the

Table 6. Percentage Overestimation of Yield Given by the Small Crop-Cut Sampling Method for Winter (Aman) Rice Crop, Bangladesh

<i>Region</i>	<i>Ten subplots, each one square meter</i>		<i>One subplot, ten square meters per field</i>	
	<i>Crop cut</i>	<i>Hill count</i>	<i>Crop cut</i>	<i>Hill count</i>
<i>Region 1</i>				
Percentage overestimate	21	13	21	19
Standard deviation	12	11	12	12
<i>Region 2a</i>				
Percentage overestimate	20	n.a.	17	n.a.
Standard deviation	17	—	14	—
<i>Region 2b</i>				
Percentage overestimate	19	9	12	6
Standard deviation	8	8	12	14

n.a. Not available.

Table 7. Variation in Bundle Weights, Nigeria

<i>Bundle</i>	<i>Funtua project area</i>		<i>Gusau project area</i>				<i>Gombe project area</i>	
	1977	1978	1976	1977	1978	1982	1977	1978
<i>Sorghum</i>								
Mean (kilograms)	46	49	26	30	32	31	27	26
Coefficient of variation (percent)	50	n.a.	n.a.	23	59	18	70	27
Number of plots	445	375	n.a.	400	462	55	474	317
<i>Millet</i>								
Mean (kilograms)	42	31	26	29	31	n.a.	24	23
Coefficient of variation (percent)	57	77	n.a.	21	52	n.a.	54	39
Number of plots	264	206	n.a.	347	346	n.a.	316	186

n.a. Not available.

Source: Agricultural Projects Monitoring, Evaluation, and Planning Unit, Federal Department of Rural Development, Nigeria.

crop should be gathered in its entirety so that the enumerator can estimate reliably the total number of units by either inspection (counting) or questioning the farmer; (d) the entire crop should also be accessible so that the enumerator can select a representative sample for estimating unit weights; and (e) when yield is required as well as production, it should be possible to associate the crop or a part thereof with a specific harvested area. There is no traditional harvest unit for root crops such as yams and cassava, and because of their bulk their tubers are handled individually.

Because of inappropriate sampling procedures or the inherent variability in traditional units, the estimates of unit weights can vary greatly from place to place and year to year. For instance, table 7 shows the variation in bundle weights for two crops at three locations over several years. Although there is a measure of consistency within each site, there are large variations between sites and in coefficients of variation over years. Some traditional units (basket, metal pail) are sufficiently uniform for this method, however, as is the ubiquitous jute sack. In that case, enumerators should be told specifically how to draw sample units for weighing, for example by selecting every tenth or fifteenth unit.

The uncertain timing of harvest complicates the enumerator's work as well as its supervision. With many types of crops, the units are removed from the field and transported to the granary or otherwise disposed of quickly; once that happens the original group of units cannot be reassembled for assessment of their total number and selection of an unbiased sample for weighing. Further complications arise if the harvesting operation is spread over several days or longer, or if the crop requires multiple harvesting. For these reasons in fact, the harvest unit method is not recommended for crops such as yams and cassava, and its utility for crops of smaller bulk (cocoyam, potatoes) must be judged according to circumstances. It is applicable for cereals, however, especially when the harvest is placed into the granaries all at the same time.

When the harvest is observed in a store it can no longer be subdivided according to the plot or field from which each part came. In general, therefore, it is recommended that the harvest unit method be used to estimate production on a per-holding rather than on a per-field or -plot basis.

Farmer Estimates of Output

In certain well-defined cropping situations, carefully obtained estimates by farmers can provide valid indications of holding production and yields. The method does not require laborious objective measurement and hence can be cheap, quick, and applicable on a larger scale compared with the procedures discussed so far. The farmer estimate method has

two special advantages in the context of monitoring and evaluation. First, because of the beneficiary context, there are good reasons to expect cooperation from the farmer. Second, the measurement of production by this method can be linked to and integrated with interview surveys of beneficiary reactions, characteristics, and so on, which cover many other topics relevant to assessing project performance. Care must be taken, however, to explain that harvest estimates are not linked to any possible food aid in order to avoid systematic underreporting. With a mobile team of enumerators the need for a heavily clustered sample can be avoided.

The method involves asking a sample of farmers to estimate their production in terms of either weight or the number of units they normally use. The survey can be either preharvest (expected output) or post-harvest. The preharvest estimate is best done plot by plot, with the enumerator and the farmer in visual contact with the growing crop. This way the enumerator can judge the validity of the farmer's response and probe for possible inconsistencies in the farmer's estimate. After harvest, the approach may be somewhat different: the estimate should be made at the farmer's house or where the crop is stored, so that where necessary the enumerator can refer to the available storage capacity as a simple cross-check.

The mean weight per unit must be estimated. In view of the variation apparent in table 7, it is recommended that, to the extent possible, the mean unit weight be determined separately for different geographic areas, crops, and seasons. The units selected for determining the mean weight should be inspected to ensure that they do not contain anything but the crop in question and are filled as is customary (with the top either leveled off or heaping to overflow). Ideally, their moisture content will be measured.

The farmer's estimates may be subject to biases which cannot be reduced by decreasing the scale of the operation and increasing the level of supervision. Inaccuracies may result from a desire to hide production from project managers or from an inherent lack of knowledge on the part of the respondent. In a well-defined situation, however, it may be possible to assess the magnitude of the biases involved by comparing farmers' reports with an appropriately matched small sample on which accurate measurements have been taken with, for example, the complete harvest method.

The reliability of the method of course depends upon the particular situation and cropping pattern. Definitive evaluation studies are lacking. In a number of countries, however, such as Bangladesh, India (Haryana State), Nigeria, the Philippines, Thailand, and Zimbabwe, encouraging results have been reported with this method. Example 13 illustrated that reporting by farmers provided much more plausible estimates of yields in the small-scale communal sector than was provided by the much

more complex crop cutting method. On the basis of such evidence and because of obvious cost and statistical advantages, a much wider use of the method is recommended. Early results from studies in five African countries confirm the validity of using farmer estimates both pre- and post-harvest.

Assessment

Information on crop production may be obtained from assessments by agricultural extension agents on the basis of their eye estimates and their general familiarity with the situation and individual farmers. The main attraction of this method is that it can be applied on a relatively large scale at only a marginal additional cost, without the need to establish a new machinery for data collection. Consequently, in certain situations with limited resources, this may be the only method available.

There are, however, two fundamental limitations to assessments. First, it is often quite unreasonable to burden extension agents with substantial responsibilities for data collection, particularly if intensive operations such as crop cutting and weighing are involved. Second, extension agents are clearly involved in the situation and cannot always be expected to observe and report the information objectively, especially when the information pertains to the quality of their own work as extension agents.

Both estimates by farmers and assessments by extension agents can also be evaluated by estimating the filled granaries' content. The granaries can even be measured and their volumes calculated if their shapes are sufficiently regular.

8 | *Exploratory Analysis*

ONCE A DATA SET has been assembled, whatever its size and complexity, the monitoring and evaluation staff must attempt to analyze it in order to disseminate information that managers can use. The scope of the required analysis should have been identified when the specifications for data collection were drawn up, and the required tabulations should have been outlined. Now the data must be reviewed and necessary revisions to the analysis program determined. Unfortunately, many monitoring and evaluation units are ill equipped for this vital stage. Many staff have received little formal training in analyzing data; others have learned to use standard statistical software packages on computers but have little understanding of the assumptions built into the forms of analysis that they select. As a result, managers often are presented either with tabulations that they are unable to digest or with misleading statements regarding significant differences revealed by the application of inappropriate statistical methods. Chapter 9 introduces some of the precautions that need to be taken in using techniques of statistical analysis, but first the need for exploratory analysis of the data must be stressed; this stage is useful in itself and provides an opportunity for those with limited training in statistics to understand the data, which in turn improves their chance of interpreting them correctly.

For monitoring purposes, simple exploratory analysis may be all that is needed. Evaluations may require more complex techniques, but exploratory analysis is still an essential first step.

Exploratory analysis seeks to reveal the simple structures and patterns in the data. We need to qualify our use of the term more specifically. A growing literature¹ on exploratory analysis of data takes this subject into formal and robust analytical techniques, which are distinguished from

1. A. S. C. Ehrenberg, *Data Reduction* (London: Wiley, 1978); D. C. Hoaglin, F. Mosteller, and J. W. Tukey, eds., *Exploring Data Tables, Trends and Shapes* (New York: Wiley, 1985); C. A. O'Muircheartaigh and C. Payne, eds., *The Analysis of Survey Data* (London: Wiley, 1977).

classical statistical analyses by their nonreliance on formal, prespecified probability models. This chapter stops short of taking the reader into the full range of such techniques. Instead it introduces the most simple initial aspects of exploratory analysis, and even then only on a selective basis, in order to demonstrate a few of the possibilities open to the analyst.

One of the main purposes of such exploratory analysis is to detect gross errors in the data which, if not dealt with before formal analysis, might lead to incorrect conclusions regarding both the range and shape of the distribution of a variable and the relation between variables. Most data sets collected through the methods described earlier will contain such errors.

Within our coverage of exploratory analysis, we include graphical inspection, ordering of data, calculation of measures of location and dispersion, detection and removal of outliers, simple transformation of the data to facilitate analysis, and assessment of the likelihood of linear relations and seasonal patterns in time series.

Simple Graphic Examination

Graphs are drawn as part of exploratory analysis to detect possible patterns that suggest lines of further analysis, not to present the results more simply. Plotting the main variables in a data set against each other or against time can give the analyst a first feel for possible relationships, make possible an assessment of the "noise" in the data, and, very important, can warn of distributional problems that may invalidate more formal analyses unless appropriate action is taken.

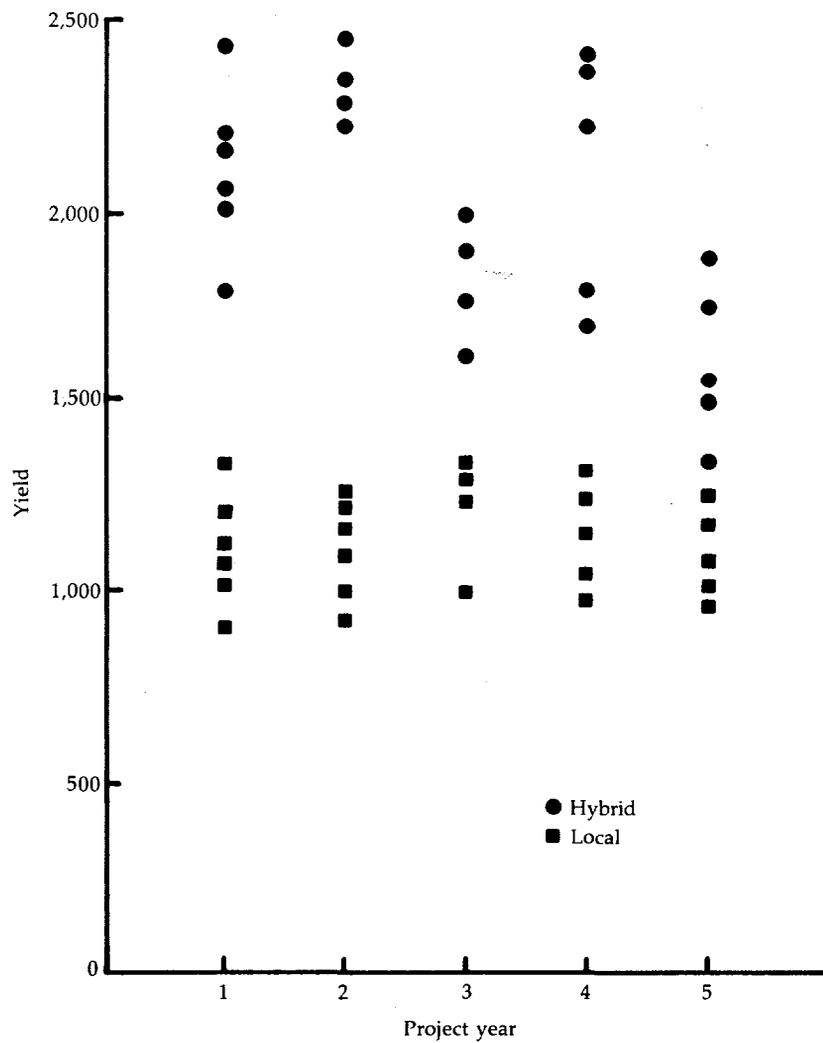
For example, consider a set of data on yields of hybrid and local maize collected annually on a sample of fields over a period of, say, five years. Figure 2 shows a scattergram of these data, a casual inspection of which shows that:

- Over the covered period, the hybrid variety yields are quite distinct from the local variety yields despite the variations within each set—the two sets belong to different "statistical populations."
- Local variety yields are essentially stable over time.
- Hybrid variety yields are much less stable, and surprisingly and worryingly, seem to decline over time.

These preliminary findings already give some useful information, and also guide analysts in deciding how to proceed. For example, it may be important to further investigate the indicated decline in hybrid yields and look for possible contributing causes.

Consider a second example, the data for which are given in table 8. These data show the amount of credit by size of holding operated by credit recipients. Is there a relation between amount of credit and size of

Figure 2. Maize Yields



holding? Regression analysis will confirm that there is. But figure 3 shows the plotted data. For holdings up to five hectares there is no relation, and for holdings above twenty hectares there is little or no relation; but a regression line for the entire data set would be as shown by the straight line. This apparent relation is misleading, for it does not hold true within either of the two distinct statistical populations of interest. In short, two distinct distributions should not be combined in one analysis. A simple plot of the data will prevent this mistake from being made.

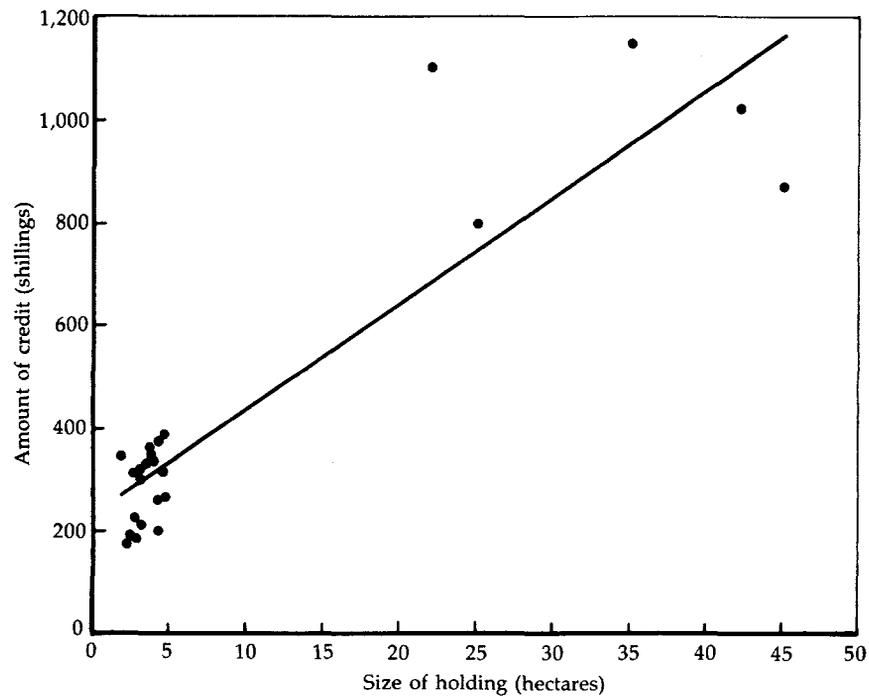
Table 8. Credit by Size of Holding

<i>Holding size (ha.)</i>	<i>Credit (sh.)</i>	<i>Holding size (ha.)</i>	<i>Credit (sh.)</i>	<i>Holding size (ha.)</i>	<i>Credit (sh.)</i>
4.1	201	2.8	225	4.2	375
22.0	1,100	42.0	1,025	4.5	317
3.0	322	3.4	330	3.8	346
4.7	266	3.0	300	45.0	875
2.2	175	1.9	349	2.6	314
2.7	186	3.1	214	2.4	192
3.7	362	25.0	800	3.9	285
4.3	263	4.6	390	3.2	340
				35.0	1,150

A final example, figure 4, shows milk sales by project year for a project with the objective of expanding such sales. Can we therefore conclude that the project veterinary and extension services were highly successful? First, consider figure 5, which shows the same milk data plotted according to the controlled price of milk as announced each year by the government, without project influence or control. We see now that milk sales are even more closely linked to price. Once again further lines of analysis are suggested. We note in passing that false correlations between two variables are very common when both are linked or confounded with the time dimension.

Ordering the Data and Measures of Location

A data set may be recorded within a data base with no easily assessable characteristics because the order of the cases may be either random or one that is not helpful in reviewing the characteristics of the distribution. If the data base is computerized, all manner of measures of central tendency and dispersion can be obtained relatively simply. But it is usually helpful first to arrange the data set in order of magnitude of the value for each important variable. The ordered distribution can be in the form of a grouped frequency distribution if the data set is large, as shown in table 9. Even a visual inspection can reveal certain important characteristics of a distribution, such as whether it is unimodal or bimodal and whether it has a long upper tail. But the ordered data also allow a quick assessment of several useful measures of central tendency, such as the median, mode, and quartiles. The arithmetic mean is the most commonly calculated measure of central tendency in a distribution, but for many of the data sets encountered by monitoring and evaluation staff the mean can be misleading—it is very sensitive to the existence of a few extreme values in the tail of the distribution. If nine households have incomes below \$200 and one household has an income of \$1,600, the arithmetic mean

Figure 3. Farm Credit Recipients

will be approximately \$300, although all but one of the cases have values substantially below this figure.

Given the popularity of the arithmetic mean, we consider first its calculation from a grouped frequency distribution. The procedure is to take the midpoint of each class interval and multiply it by the number of cases within the class, sum the result of this calculation over all classes, and divide by the total number of cases. In algebraic notation we denote it as

$$\bar{x} = \frac{\sum_{i=1}^N f_i x_i}{\sum_{i=1}^N f_i}$$

where N is the number of classes, x_i the midpoint of class i , and f_i the number of cases in class i .

If we attempt to calculate the mean income from the data in table 9, we see the problem of the lack of specified limits for the lowest class (100 or

Figure 4. Milk Sales by Year

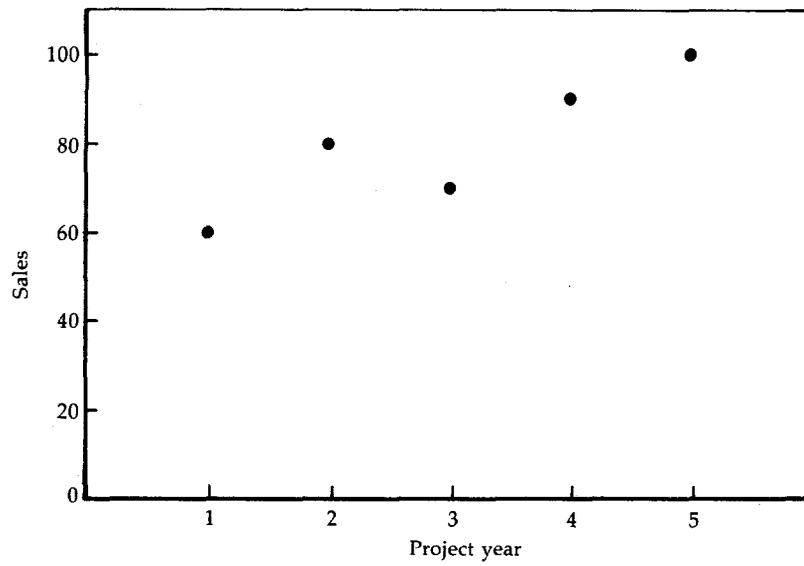


Figure 5. Milk Sales by Price

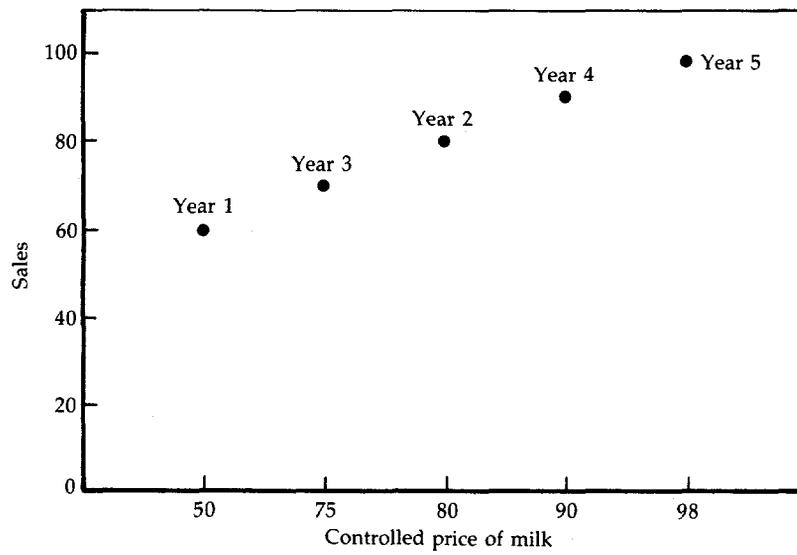


Table 9. Distribution of Households by Nine Income Classes

<i>Income class</i>	<i>Number of households</i>
100 or less	32
101–200	81
201–300	119
301–400	143
401–500	162
501–600	67
601–1,000	86
1,001–2,000	42
More than 2,000	18
<i>Total</i>	750

less) and the highest class (more than 2,000). Because of this, the calculation of the midpoint of these classes will be approximate at best: we may take 50 as the midpoint of the lower class (the real lower limit is zero) and 3,000 as the midpoint of the upper if examination of the data shows that virtually no household incomes exceed 4,000. Then the estimate of the mean is

$$\frac{(50 \times 32) + (150 \times 81) + (250 \times 119) + \dots + (3,000 \times 18)}{750}$$

$$= 518.8$$

An open-ended upper class does not affect the median or mode, measures that are relatively unaffected by extreme values.

The mode is the value or frequency class that occurs most often. In table 9 it is the frequency class 401–500. If there are two distinct peaks in a distribution, it is said to be bimodal: this may be an important finding in itself. For example, bimodality in monthly rainfall data reveals the existence of two wet seasons. When dealing with grouped frequency distributions, the modal class depends on the choice of class boundaries. Table 10 presents the same data as table 9 but with different class intervals. The modal class is now 201–400, very different from that identified in table 9. Strictly, the mode is only useful when the distribution is given in narrow, equal-class intervals.

The median—more useful than the mode, and often more useful than the arithmetic mean—is the value in the middle of distribution. There are as many cases with values below the median as cases with values above it. If, say, 101 cases are listed in ascending order, the median will be the value of case 51 (50 values are less than this and 50 values exceed it). In percentage terms, the median is the value of the 50 percent point in the distribution. If the number of cases is even, say, 100, the median lies half-

Table 10. Distribution of Households by Five Income Classes

<i>Income class</i>	<i>Number of households</i>
200 or less	113
201-400	262
401-600	229
601-1,000	86
More than 1,000	60
<i>Total</i>	<i>750</i>

way between the fiftieth and fifty-first ordered values. To generalize, if the number of cases, n , is odd, the median is the value of the $(n + 1)/2$ case. If n is even, the median is the average of the values of the $n/2$ and the $(n/2) + 1$ cases.

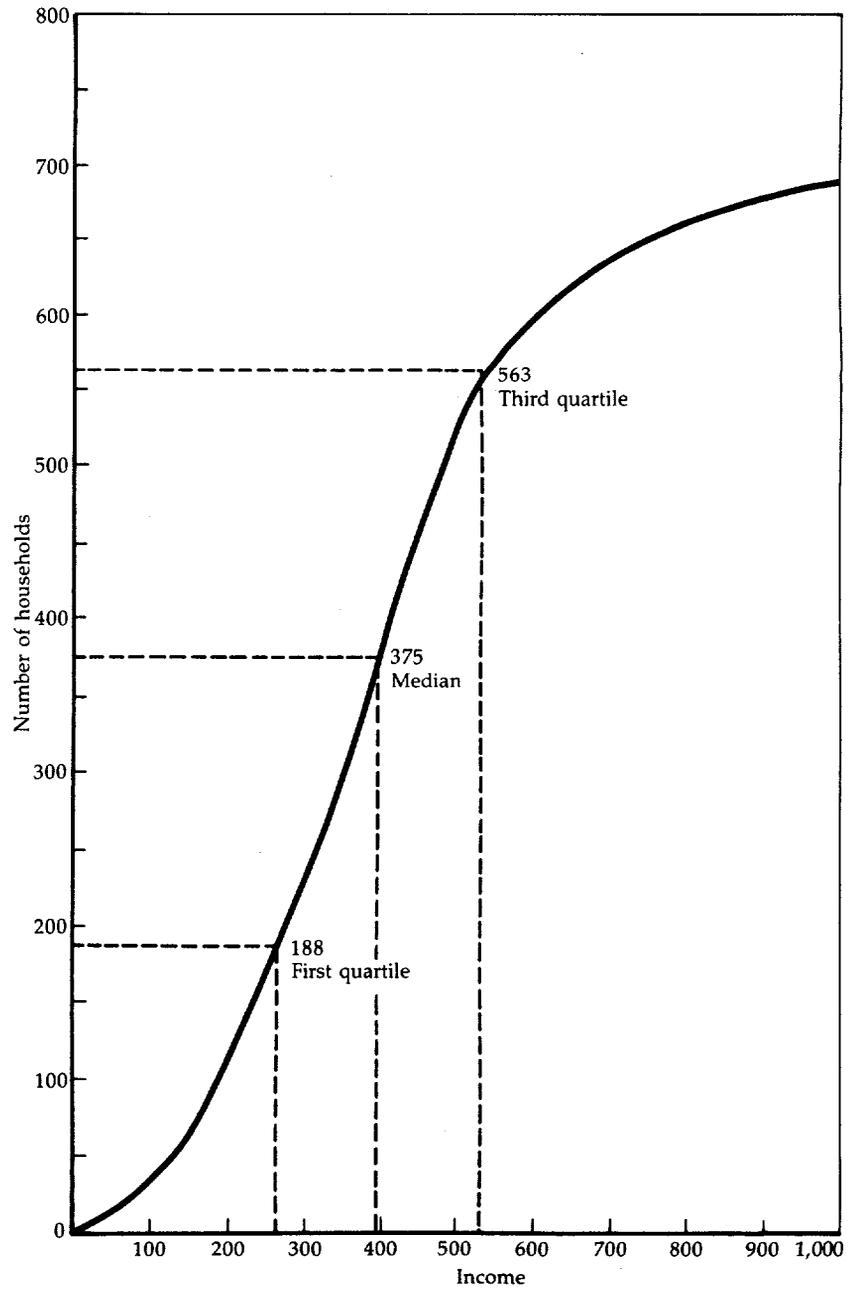
If we are dealing with a grouped frequency distribution, it is useful to plot the graph of the cumulative frequency and read off the median as the point where the cumulative frequency reaches 50 percent of the overall total. Table 11 presents the data of table 9 with the cumulative totals added. Figure 6 shows the partial graph of this cumulative distribution with the median identified. Using the same approach, the lower (25 percent point) and upper (75 percent point) quartiles can be identified.

Most data sets examined in small farmer development projects are likely to contain some major errors that were undetectable in the editing and checking process. Misreporting or misobservation by one or more enumerators is probable. In some cases misreportings result in values in the data set which are "outliers"—they are well above or below the values observed for most of the cases. When dealing with such data sets, the me-

Table 11. Cumulative Distribution of Households by Nine Income Classes

<i>Income class</i>	<i>Number of households</i>	<i>Cumulative number of households</i>
100 or less	32	32
101-200	81	113
201-300	119	232
301-400	143	375
401-500	162	537
501-600	67	604
601-1,000	86	690
1,001-2,000	42	732
More than 2,000	18	750

Figure 6. Cumulative Frequency Distribution



dian is more useful than the arithmetic mean at the stage of exploratory analysis.

Trimming and Transforming Data

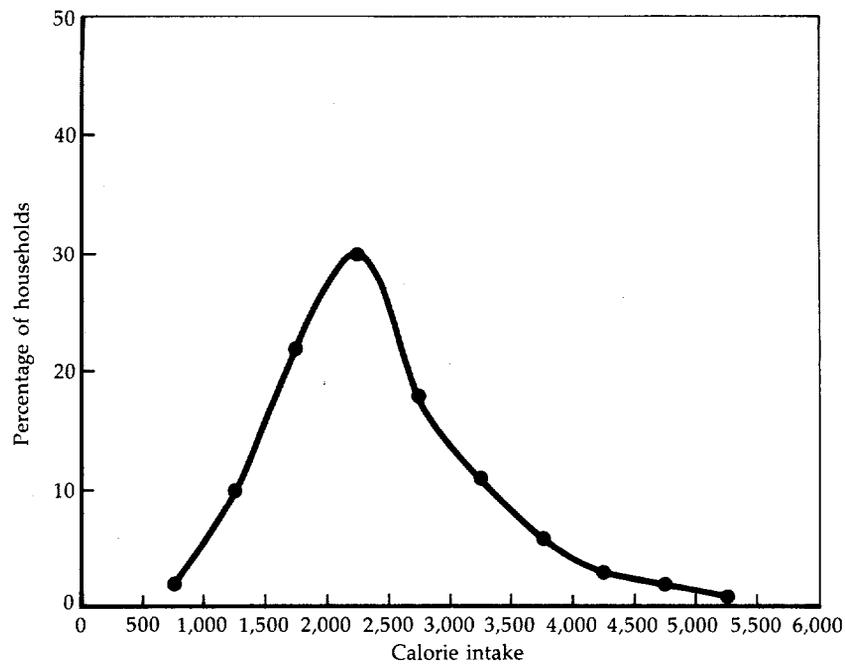
Following preliminary graphic and ordered distribution examination of the data, some adjustments may be indicated to remove extreme outliers and to facilitate further interpretation of the data. Although the median can be used in place of the mean to reduce the disturbing influence of outliers, the mean will likely retain its prominence with most analysts in the more formal stages of analysis and in presentation of the results. Removal of outliers may be a necessary step if such is the case. The decision to remove an observation from the data set must be carried out according to an agreed rule, for otherwise the analyst opens himself to charges of improper manipulation of the data. The graphic review stage will have revealed one type of outlier: values that do not belong to the same statistical distribution as the others. The main body of the data lies within a certain range, then there is a break, then there are the outliers. Figure 3 shows such a case, in which the credit values for the few larger holdings are in a distinctly different range from the data's main body. This technique can also reveal the presence of values that have been recorded with gross errors. Outliers in other data sets are harder to identify. A distribution of calorie intakes of households may be continuous but have a very extended upper tail of the distribution, as shown in figure 7. Depending on the variable, there may be a value such that any observation above it is of dubious validity. It may be considered that any measured calorie intake above 5,000 is suspect and to trim the data set at this point will be eliminating cases with higher values. Again we stress that such practices must be done according to rules agreed on before the data are examined.

Transformation of the data can be very simple if it is to facilitate simple review of the data or quite complex if it is to remove skewness in the data and induce what is known as the normal distribution or promote a linear relationship between two variables. Transforming temperature readings from Fahrenheit scale to the Celsius (centigrade) scale so that zero corresponds to the freezing point of water is an example of the first simple type of transformation. Logarithmic and exponential transformations, often used by econometricians when handling income data, are examples of the latter.

At the exploratory stage, we are interested in simple transformations, such as the examples that follow.

Lateral Shift of the Scale

Sometimes a simple change of scale assists the analyst at the exploratory stage to get a feel for the data set. Table 12 shows a set of data of the num-

Figure 7. Distribution of Households by Individual Calorie Intake

ber of credits approved by various credit branches. Column 2 shows the data with 1,000 removed from each value. Alternatively, the scale can be shifted by expressing each value as a difference from the mean of the data set (1,265), as shown by column 3.

Conversion to Percentages

A set of data that reveals little may become very informative when the values are converted into a percentage of some standard or target. Suppose, for the data in table 12, that the respective branches of the credit bank had each been set a target of 1,200 loans for the given period. The data could be shown as in column 4.

Standardized Index

A percentage is one type of index. Another popular transformation is to take the mean away from each value and divide the result by the mean, producing a standardized value given by $x'_i = (x_i - \bar{x}) / \bar{x}$. This index is useful for comparing two data sets with different means, as is shown in

Table 12. Variations of Transformed Data

Region A credit branch	Number of credits approved (1)	Credits issued in excess of 1,000 (2)	$x - \bar{x}$ (3)	Percentage of target of 1,200 (4)
1	1,030	30	-235	86
2	1,250	250	-15	104
3	1,170	170	-95	98
4	1,330	330	+65	111
5	1,290	290	+25	108
6	1,480	480	+215	123
7	1,180	180	-85	98
8	1,320	320	+55	110
9	1,460	460	+195	122
10	1,140	140	-125	95

table 13. It clarifies the relative size and shape of the variation around the mean in the two data sets.

A more sophisticated transformation of this type converts the data into what are known as standardized normal deviates, where the denominator is a measure of the variability in the data set (which is explained below): $x'_i = (x_i - \bar{x}) / s_x$.

The resulting distribution will have a mean of zero and a standard deviation of one. Its plot will show the shape of the variability around the mean, removing the effect of the actual size of the mean and the magnitude of the overall variability. It can then be compared with a normal dis-

Table 13. Data Converted to Standardized Index

	Credit branch		Standardized index		
	Region A ^a	Region B ^b	Region A	Region B	
1	1,030	11	2,020	-19	-19
2	1,250	12	2,450	-1	-1
3	1,170	13	2,710	-8	+9
4	1,330	14	2,100	+5	-15
5	1,290	15	2,850	+2	+14
6	1,480	16	2,280	+17	-8
7	1,180	17	2,170	-7	-12
8	1,320	18	2,640	+4	+6
9	1,460	19	2,870	+15	+15
10	1,140	20	2,780	-10	+12

a. Mean of region A = 1,265.

b. Mean of region B = 2,487.

tribution, which is an important standard that is assumed in a great deal of statistical analysis.

Measures of Dispersion

The simplest measure of the extent of dispersion of the values in the data set is the range, the difference between the highest and lowest values. It is not a satisfactory measure for a skewed distribution with a long upper tail because it is determined by the highest value—one extreme outlier will dramatically increase the range.

The commonest measure of dispersion is the standard deviation of the distribution. This is defined as the square root of the variance where

$$\text{Variance} = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$$

Each case is expressed in terms of the deviation from the mean; these deviations are squared and summed, and the results are divided by the number of cases (N).

If the variance is to be estimated from a sample n drawn from a larger population N , the divisor in the above equation is $n - 1$, not n , that is

$$\text{Variance} = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{n - 1}$$

and the estimated standard deviation is then denoted by

$$s_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n - 1}}$$

If the data are in the form of a grouped frequency distribution, we have the weighted estimate of s_x , similar to the weighted arithmetic mean given earlier:

$$s_x = \sqrt{\frac{\sum f_i (x_i - \bar{x})^2}{\sum (f_i) - 1}}$$

where f_i is the number of cases in class i . The reader may care to try this calculation using the data in table 9.

The arithmetic mean and the standard deviation are the first two moments of a distribution. One can compute the third and fourth moments that describe the extent of skewness (distortion from the symmetrical) and kurtosis (whether the distribution is flat or peaked), but these have few practical applications in monitoring and evaluation. If a distribution

is normally distributed (the classic bell-shaped symmetrical curve, for which mean = median, no skewness, no kurtosis, no extreme outliers), the mean and standard deviation define the distribution.

Fitting Straight Lines

Preliminary graphical examination of the data may reveal the existence of an approximately linear relationship between two variables. At the exploratory stage of the analysis this may be sufficient, but the analyst may wish to quickly draw in the regression line, or to have an approximate idea of the slope of the line. The fitting of a straight line of the form $y = a + bx$ to a data set $(y_i, x_i), i = 1, \dots, n$, is one of the basic statistical calculations taught at an early stage of statistical tuition and uses the method known as least-squares regression. There are quicker methods (and given that both variables are subject to error some of these are more robust than that of least squares), of which we recommend the following.

Having ordered the data according to the ascending values of x (see above), calculate the means of the variables for the lower quartile and upper quartile respectively. Plot these points, and draw the regression line through them. In notational form, the estimate of the slope of this line is given by $\hat{b} = (\bar{y}_H - \bar{y}_L) / (\bar{x}_H - \bar{x}_L)$, where \bar{y}_H and \bar{x}_H are the means of the upper quartile cases and \bar{y}_L and \bar{x}_L are the means of the lower quartile cases. This is very quick to do and is a reliable estimator of the true linear relationship. An example is shown in figure 8.

Analysis of Residuals and First Differences

At the exploratory analysis stage, the analyst forms some hypotheses concerning the underlying relations in the data set. Let us assume that one such hypothesis is that there is a linear relation between two variables in the data set. A study of the residuals, defined as the differences between the original values and the estimates for them made on the basis of the fitted relation, will enable the analyst to test the validity of the relation. An informal test of the validity of the relation is the absence of patterns in the residuals. Alternatively, if a pattern emerges when the residuals are plotted against the fitted value, the type of pattern may suggest how the model of the relation can be improved.

Consider the example shown in table 14. The hypothesized linear relationship has been calculated by the quartile method described above. The third column shows the residuals from the fitted line. Figure 9 shows the plot of the residuals against x . The figure shows very clearly that the spread of the residuals increases as x increases. This phenomenon is a strong indicator that the variance of the dependent variable (y) increases for higher values of the independent variable. This is important, for any

Figure 8. Example of Regression Line Fitted by Quartile Method

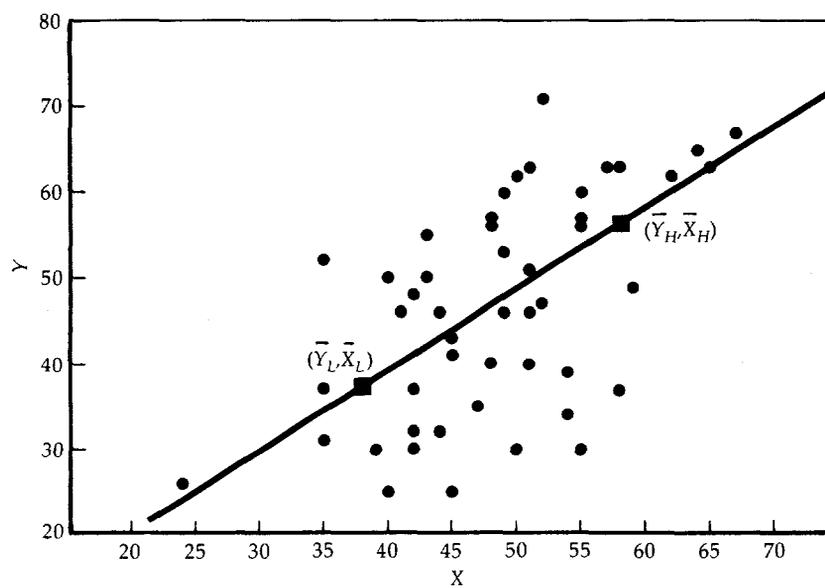
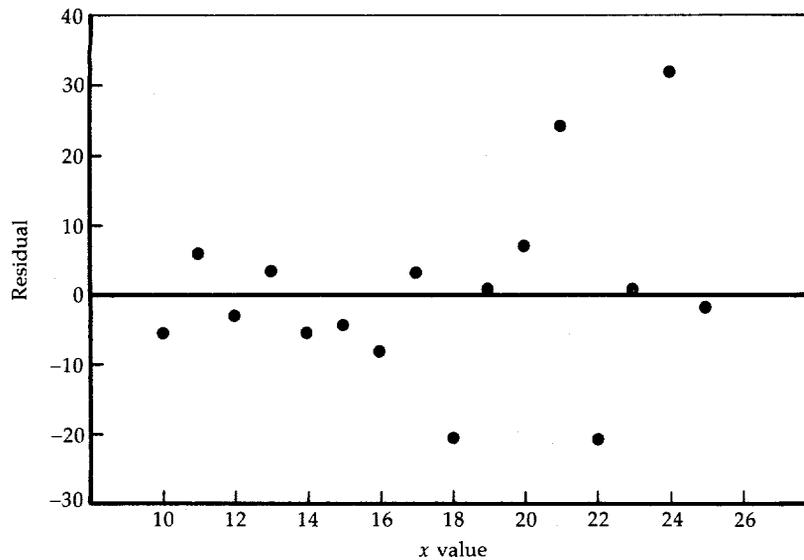


Table 14. Residuals from Fitted Linear Relationship

<i>x</i>	<i>y</i>	<i>Residual</i>
10	50	-5.5
11	65	+5.8
12	60	-3.0
13	70	+3.3
14	65	-5.5
15	70	-4.3
16	70	-8.0
17	85	+3.3
18	65	-20.5
19	90	+0.8
20	100	+7.0
21	120	+24.3
22	80	-20.5
23	95	+0.8
24	140	+32.0
25	110	-1.8

Note: Fitted line is of form $y = a + bx$; $\hat{b} = (\bar{y}_H - \bar{y}_L) / (\bar{x}_H - \bar{x}_L) = 3.75$,
 $\hat{a} = \bar{y} - 3.75\bar{x} = 17.8$.

**Figure 9. Residuals Plotted against Their x -Value
(Demonstrating Variance Increasing with x)**



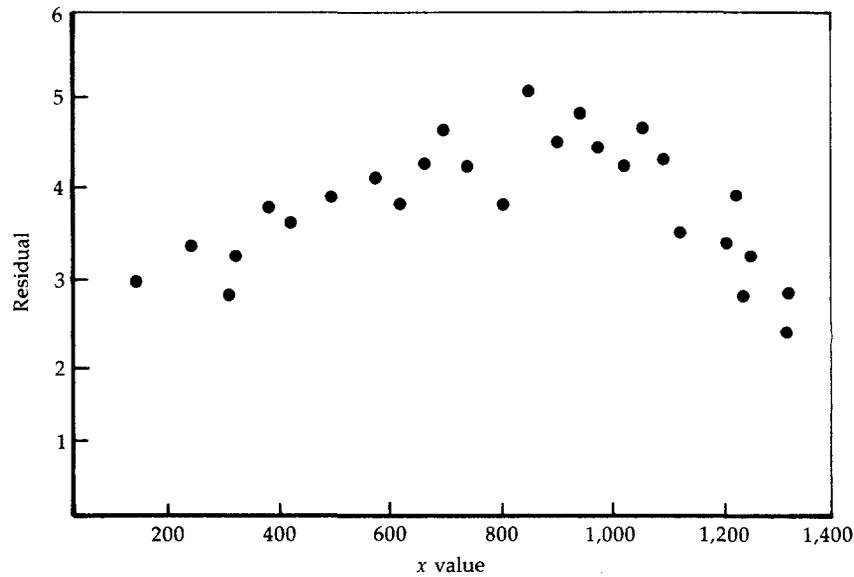
later application of formal least-squares linear regression analysis has as one of its conditions the requirement that the variance of y is constant across the range of x . Transformation of the data is indicated. The other common result of plotting the residuals is that they form a curved band, as shown in figure 10 for a hypothetical distribution. This is an indicator that the true relation between x and y is nonlinear.

We have referred to the possibility of attributing a spurious relation between two variables because of the confounding effect of a third variable, such as time. A high correlation between two variables when each pair of observations has been made at successive time intervals (for example, consecutive seasons) may be caused by the fact that both variables are correlated with time. This can be tested by seeing if the relation is maintained when the two series of first differences (that is, period-to-period changes in both variables) are examined. If the first differences are also highly correlated, confidence that there is a real relationship between the two variables is increased.

Moving Averages

An important monitoring indicator may be the movement of prices of certain foods or commodities in local markets. Weekly wholesale and retail prices of food crops from a sample of markets in the project area will

**Figure 10. Residuals Plotted against Their x -Value
(Demonstrating Nonlinearity of x - y Relationship)**



be of interest to project managers. Over an extended period of time, a series of data exhibit three characteristics: an underlying trend, cyclical or seasonal oscillations, and random movements. With a long data series recorded at narrow intervals, a simple smoothing to eliminate the random noise is required; the use of moving averages may serve this purpose.

Consider a series of prices c_1, \dots, c_n . A five-point moving average produces a new set of points:

$$\begin{aligned}\bar{c}_3 &= \frac{c_1 + c_2 + c_3 + c_4 + c_5}{5} \\ \bar{c}_4 &= \frac{c_2 + c_3 + c_4 + c_5 + c_6}{5} \\ &\vdots \\ \bar{c}_{n-2} &= \frac{c_{n-4} + c_{n-3} + c_{n-2} + c_{n-1} + c_n}{5}\end{aligned}$$

When plotted, these moving averages will reveal the cycles and trends but remove most of the random movements. The longer the period chosen for the moving average—three-point, five-point, seven-point, and so on—the more the short-term movements are suppressed. The moving average data points are fewer than the original series—points are “lost” at each end.

Proportions

In chapter 4 of the companion volume and throughout this volume, we have stressed the importance of adoption rates in various forms. Estimating the mean of a proportion or rate is of course a relatively simple calculation, as is the estimation of its reliability from a sample—as discussed in chapter 6 of this volume. The adoption rate is expressed as the number who are observed to adopt divided by the total number of observations. If a simple random sample is used, the variance is given by the simple formula $\text{var}(p) = p(1-p)/(n-1)$, where p is the observed adoption rate. Apart from their intrinsic usefulness, therefore, analyses based on rates are appealing in that the calculations really do require nothing more than a pencil and paper.

Analysis of Subsamples

We have not recommended collecting data from very large samples nor including all variables that may be of some possible relevance. Our keynote has been to collect the minimum data that are essential. Nevertheless, analysts may need to review a data set of moderate size and complexity, particularly when the data refer to a time series accumulated over several years. The steps of exploratory analysis we have introduced in this chapter can be facilitated further by first randomly selecting a subsample of the cases in the data set and using this subsample for a preliminary examination of possible features of interest which can be tested further with the use of the full set. If several hundred cases are available, a preliminary selection of, say, 100 may serve at the stage of exploratory analysis.

9 | *Statistical Analysis of Data*

THE EXPLORATORY OR INITIAL ANALYSIS of data will suffice in many instances. Suitable tabular and graphic presentation with text, as discussed in chapter 10, is the next and only required stage of the process of turning the data into information. In some instances, however, a more formal statistical analysis will be needed. There is of course no clear dividing line between exploratory analysis and full formal analysis—some of the simple techniques discussed in chapter 8 are perfectly valid techniques in their own right. This chapter introduces the reader to more advanced analytical tools, which complement the simpler ones. We were selective in choosing these tools and of necessity are brief in describing them. Powerful as they are, they can be misused and misinterpreted even by professional statisticians; if used by those not fully versed in the underlying theory, they can be dangerous indeed.

The advent of microcomputers and statistical software packages has made it possible to do significance testing, correlation and regression analysis, and analysis of variance merely by following a few instructions to summon the applicable part of the software program. Unfortunately, the use of these packages does not normally require the operator to demonstrate knowledge of the underlying assumptions concerning the structure of the data which must hold true if the analysis is to be valid. An eminent statistician said recently, "Perhaps a warning message should be built into subset regression programmes. My guess is that, were this to be done, users would obtain the warning more often than they would obtain any actual answers. It has been said, 'If you torture the data for long enough, in the end they will confess.' The data will always confess, and the confession will usually be wrong."¹ This chapter starts by comparing two sample means; proceeds to consider 2×2 tables; and then by extension moves to larger cross-tabulations, to comparison of multiple means

1. A. J. Miller, "Selection of Subsets of Regression Variables (discussion)." *Journal of the Royal Statistical Society*, series A, vol. 147, no. 3 (1984):410–15.

through the use of analysis of variance, and finally to correlation and regression.

Comparing Two Sample Means

In simple applications of statistical analysis to project monitoring, we may want to compare the mean of a variable achieved in one zone with that achieved in another; or to compare the average rate of adoption achieved by project participants with that of nonparticipants. We may want to know if these means estimated from sample observations are significantly different, that is, that the difference is not due to the particular samples that were selected.

We need to introduce here an extension of the concept of standard deviation as a measure of dispersion, which was introduced in chapter 8—namely, the standard deviation of all possible sample estimates for a given sample size of a mean, percentage, or proportion. This is called the standard error of the sample estimate, and it was presented for various sample designs in chapter 6. For illustrative purposes, we consider the standard error of a mean when calculated from the data collected by using a simple random sample of size n .

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} = \sqrt{\frac{\sum(x_i - \bar{x})^2}{n(n-1)}}$$

From here on, we shall refer to the standard error of a mean, which implies that we are dealing with sample data and estimating the value from the sample itself.

When comparing means and their standard errors, one further measure is useful: the coefficient of variation. Clearly, a standard error of 2 has a different application in terms of a mean of 10 than in terms of a mean of 100, so we convert the standard error into relative terms: coefficient of variation = $V_{\bar{x}} = s_{\bar{x}}/\bar{x}$.

Now consider a case where the application rate of fertilizer has been estimated for each of two project zones using a random sample of twenty farmers in each zone. The means and standard errors are:

Zone 1: mean application rate, 110 kilograms; standard error, 9.4

Zone 2: mean application rate, 160 kilograms; standard error, 19.5.

Do we have sufficient evidence that the difference in application rates in these zones is not due to the chance involved in the selection of farmers? The test used to determine this is known as the t -test. We calculate a value of t given by

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{(s_{\bar{x}_2}^2 + s_{\bar{x}_1}^2)}}$$

Using the values for the zones as shown above, we find that $t = 2.31$.

This well-known distribution is quantified in standard statistical tables. The extra number we need in order to use these tables is the number of degrees of freedom. In our example it is given by $(n_1 + n_2 - 2) = 38$. We find from standard tables that the 95 percent point of the t -distribution with 38 degrees of freedom equates to a value of $t = 2.02$. Because our computed value is higher than this, we can say with 95 percent confidence that farmers in zone 2 tend to use higher application rates than farmers in zone 1.

If t tables are not available, do not despair. Except for very small samples, the t -distribution is almost identical to the normal distribution. But tables for this may not be available either. Then take as a rule of thumb that if the size of the samples exceeds 10, the value of t needs to be greater than 2 if you are to assume that there is a significant difference between the means.

2 × 2 Tables

We move from a comparison of two means to a slight extension where the data are assembled in a 2×2 table for the purpose of examining the existence of associations and interrelations. The use of a 2×2 table is most easily demonstrated by using an example (table 15) to demonstrate the simple calculations that can be made. To generalize the results we will make use of table 16, which denotes each cell by an algebraic symbol.

The difference in the proportions "adopting" for project and non-project farmers gives an idea of project performance. The differences in proportions are as follows. From table 15:

$$\begin{aligned} p_1 - p_2 &= \frac{180}{250} - \frac{60}{150} \\ &= 0.32 \end{aligned}$$

From table 16:

$$p_1 - p_2 = \frac{a}{a+b} - \frac{c}{c+d} = \frac{ad - bc}{(a+b)(c+d)}$$

Table 15. Adoption of Practice by Project and Nonproject Farmers

<i>Population</i>	<i>Adopt</i>	<i>Do not adopt</i>	<i>Total</i>
Project farmers	180	70	250
Nonproject farmers	60	90	150
<i>Total</i>	240	160	400

Table 16. Table 15 in General Form

Item	Yes (1)	No (0)	Total
In (1)	a	b	$(a + b)$
Out (0)	c	d	$(c + d)$
Total	$(a + c)$	$(b + d)$	$n = (a + b + c + d)$

If several tables are being studied, comparisons of p_1 's and p_2 's can be made.

For making comparisons across a number of tables, a measure that ranges from -1 to $+1$ may be more convenient. One such is Yule's Q , which is calculated as follows.

From Table 15:

$$Q = \frac{180 \times 90 - 60 \times 70}{180 \times 90 + 60 \times 70}$$

$$= 0.59$$

From Table 16:

$$Q = \frac{ad - bc}{ad + bc}$$

Q was originally devised as a measure of association and is appropriate here because a difference in the odds on adoption for project and nonproject farmers is equivalent to an association between "being in the project" and "adopting the practice." Q takes values ranging from 0 (no association) to $+1$ in one direction (complete positive association) and -1 in the other (complete negative association). The larger the figure, the greater the strength of the association it summarizes.

Another measure sometimes used is the correlation coefficient r (which also takes values lying in the range of -1 to $+1$). We discuss the general form for r later in this chapter, but in the particular case of a 2×2 table r can be expressed as follows (using the assignment of the numeric codes 0 and 1 in table 16).

From table 15:

$$r = \frac{12,000}{\sqrt{(250 \times 150 \times 240 \times 160)}}$$

$$= 0.316$$

From table 16:

$$r = \frac{ad - bc}{\sqrt{[(a + b)(c + d)(a + c)(b + d)]}}$$

The result for r will be of the same sign as Q (because the numerator is the same, the positive root is always taken in the denominator of r) and r will always be closer to 0 than Q . It is often recommended that the value of r^2 be considered, rather than the value of r , since r^2 equals the proportion of the variation explained. The value of r^2 in this case is equal to only 0.1.

The quantity $(ad - bc)$ occurs again in another often-used test of association, termed the chi-square test. Suppose there is in reality no association between the two variables. In this case, it would then be expected that the $(a + b)$ farmers, the $(c + d)$ farmers, and the total of farmers would all divide in the same proportion. In the cell in which a is found, the number would be expected to satisfy

$$\frac{a}{a + b} = \frac{a + c}{n}$$

giving an expected value in the first cell under an assumption of no association of

$$\frac{(a + b)(a + c)}{n}$$

The number actually observed is a , and so the difference (D_{11}) between the observed and "expected" values is as follows. From table 15:

$$\begin{aligned} D_{11} &= \frac{(180 \times 90) - (70 \times 60)}{400} \\ &= 30 \end{aligned}$$

From table 16:

$$\begin{aligned} D_{11} &= a - \frac{(a + b)(a + c)}{n} \\ &= \frac{ad - bc}{n} \end{aligned}$$

Keeping the marginal totals constant, it will be seen that $D_{11} = D_{22} = -D_{12} = -D_{21}$.

Clearly, the larger $(ad - bc)/n$, the bigger is the difference between the observed value and that expected on the assumption of no association, and the stronger is the evidence for the alternative view that there is a relation between the variables.

Using this feature, the chi-square value is calculated as follows. From table 15:

$$\begin{aligned} \chi^2 &= \frac{400 (12,000)^2}{250 \times 240 \times 160 \times 150} \\ &= 400 (0.316)^2 \end{aligned}$$

From table 16:

$$\chi^2 = \frac{n(ad - bc)^2}{(a + b)(a + c)(b + d)(c + d)}$$

$$\begin{aligned} &\text{or substituting } r \text{ from above,} \\ &= nr^2 \end{aligned}$$

If there were no association between the variables, samples giving a value of χ^2 greater than 1 would occur, on average, approximately once in three samples, and values greater than 3.8 would occur only once in twenty times. One reason for calculating this particular quantity is that it is a convenient basis for combining results from a number of similar sample tables. If the association in all, say k , tables is in the same direction (that is, if $[ad - bc]$ is the same sign in all k tables), then the sum of χ^2 s from the tables can be added together and tested at the desired level against the value of χ^2 with k degrees of freedom (details in most sets of statistical tables).

Single-figure measures provide only a first stage of analysis, and the reduction of the data into 2×2 tables may be an oversimplification. Moreover, the description of these measures here is introductory only. The effects of sampling design features, for example clustering, have not been dealt with. But an analysis of this kind can be done quickly without computing aids and is quite sound in testing for possible relationships.

Cross-Tabulations

Not all tables which show data from the information system can be reduced to a simple 2×2 format. Two-way tables or cross-tabulations, in which one variable is distributed across the columns and the other down the rows, are probably the commonest form of data presentation. It is a powerful way of communicating information, and the presentational aspects are discussed in the next chapter. This chapter touches on the analysis of a cross-tabulation.

In one common form of cross-tabulation, each observed case is allocated to a cell of a table depending on its categorization according to the two dimensions of the table. Table 17 shows an example in which farmers were asked to assess the quality of the extension service. Their assessments have been tabulated according to the size of farm of each respondent. One of the tests that we may wish to perform is one about whether the tendency of farmers to rate the service as satisfactory is related to farm size. For this we extend the application of the chi-square test introduced earlier. We calculate the χ^2 value as

$$\chi^2 = \sum \frac{(\text{observed value} - \text{expected value})^2}{\text{expected value}}$$

Table 17. Rating of an Extension Service

Rating	Farm size (ha.)					Total
	0-5	5-10	10-15	15-20	20+	
Very good	9	14	15	15	8	61
Good	36	41	14	7	6	104
Fair	19	15	4	3	1	42
Poor	12	7	2	2		23
Total	76	77	35	27	15	230

The expected value for the top left-hand cell (very good / 0-5 hectares) is obtained by multiplying the row and column total in which the cell lies and dividing by the overall total. The expected total for very good / 0-5 = $76 \times 61 / 230 = 20$ and expected value for good / 0-5 = $76 \times 104 / 230 = 34$ and so on.

$$\begin{aligned}\chi^2 &= \frac{(9 - 20)^2}{20} + \frac{(36 - 34)^2}{34} + \dots + \frac{(0 - 1)^2}{1} \\ &= 37\end{aligned}$$

The degrees of freedom associated with this test are given by (number of rows - 1)(number of columns - 1) = 12.

Reference to tables of the chi-square distribution with this number of degrees of freedom shows that the probability of such a high value for χ^2 if there was no association between assessment and size of farm is very small (less than one in a thousand).

Another popular form of cross-tabulation does not use a number of cases as the cell entry, but rather a mean, percentage, or total that has been calculated from the values for the number of cases in the cell. An example is shown in table 18.

Each entry, although a percentage, is a percentage of a cell total that is not shown. The percentages therefore are independent of each other and do not sum to 100 either vertically or horizontally. Mistakes are commonly made in interpreting such tables and in drawing conclusions. Moreover, if the numbers in each cell are not known, very little further analysis can be attempted. We shall say more on this in the next chapter.

Comparing Differences from Multiple Groups: The Analysis of Variance

Earlier in this chapter we introduced the *t*-test for testing the significance of the difference between two means. In most cases, as with table 18, the problem facing the analyst is to test for significant differences between a

Table 18. Proportion of Users of Hired Labor
(percent)

Zone	Farm size (ha.)			Total
	0-5	5-10	10+	
A	21	67	86	52
B	12	79	90	54
C	27	71	83	46
D	7	59	68	33
Total	15	68	77	47

multiple set of means. A typical question the analyst addresses is whether there are significant differences in response between the various zones. The null hypothesis is that there is no such difference. The analysis aims to test this hypothesis; if it is found to be unlikely it is concluded that farmers in different zones react differently. We could conduct a series of *t*-tests comparing each mean with every other mean. For two reasons, this is not the way to proceed: the significance tests will be incorrect, and there is a better way.

Suppose we make several comparisons of means using the 95 percent confidence level of the *t*-distribution. In any single test, therefore, there is a 0.05 chance of falsely concluding that there is a significant difference, when in fact there is none. If several *t*-tests are performed, the probability that at least one of these results in an incorrect conclusion of a significant difference increases rapidly. The probability is 0.23 for five tests and 0.40 for ten tests.

We have seen analyses produced by computer statistical software packages that provide a long series of *t*-values, which have been reviewed by the analyst, who has then chosen the significant differences based on them. From the above, it follows that there is a good chance that some of these choices were made incorrectly.

The technique used to test multiple means is termed the analysis of variance. This principal area of formal statistical analysis is covered extensively in standard texts; therefore, we do not go into the theory underpinning this analysis here, but we offer a brief explanation of its function and a warning regarding its misapplication, and in some cases limited usefulness.

Until recently there was not much chance that an analyst would carry out an analysis of variance without fully understanding its underlying assumptions. Ease of access to computers and statistical software packages has changed the situation, however. The analysis of variance routine in statistical packages can be accessed with only a few control instructions in the software manual. What is actually being calculated need not

be understood. The danger of drawing erroneous conclusions in such circumstances is high.

Consider a simple case of a data set consisting of, say, crop yields obtained in fields contained in a number of zones. This data set is variable—values range over a certain set of values. We want to assess whether the location of the fields by zone is a principal contributor to explaining this variation or whether the variation may be described as caused by chance fluctuations from field to field irrespective of the zonal location. In other words, we consider the total variation around the mean of our data set and apportion it according to the variation *within* each zone around the mean of each zone, and the variation *between* the means of the zones. We compare the latter between-zone variance with the former within-zone variance. If this ratio is above a certain figure, we conclude that the zone is an important cause of differences in yields.

This brief statement covers the basic principle of analysis of variance. The actual calculations need to be carried out according to precise rules, and when we are dealing with two or three possible explanatory variables—say zone, farm size, and season—the calculations become quite complex, although well within the capacity of the aforementioned computer software packages.

In its simplest form, the model we have used is of the form $x_{ij} = \bar{x} - \varepsilon_{ij}$, that is, that the yield obtained by the j^{th} farmer in the i^{th} zone is equal to the overall mean yield plus or minus a random number, denoted by ε_{ij} , which may have a negative, positive, or zero value for any one farmer.

Several assumptions are made about ε_{ij} . They must be independent of each other, must be distributed around a zero mean and have a common variance, and must be normally distributed for the F -test to be valid. In many cases these assumptions do not hold true. Sometimes the requirements can be met by suitable transformation of the data, but most software packages do not warn the uninitiated that such transformations are required. The computer will conduct the calculations and provide the summary regardless. The summary output is in the form of an F -value (equivalent, in a way, to the t -value discussed earlier in the case of comparing two means).

What are the consequences of using tests such as the t -test and F -test when the data do not conform to the model that underpins the validity of these tests? We quote a standard work²:

Usually, but not invariably, the true significance probability is larger than the apparent one; that is, too many significant results are obtained. . . . The most serious disturbances appear to arise when the ex-

2. William G. Cochran and Gertrude M. Cox, *Experimental Designs* (New York: Wiley, 1957), p. 91.

perimental error variance is not constant over all observations . . . the use of the same estimated variance for both comparisons would lead to *t*-tests that were completely erroneous . . . a transformation [of the data may] place the data on a scale on which the error variance is more nearly constant.

For routine testing to give first indications of differences, these problems may not be of great concern in many evaluation applications. But they must be borne in mind if very formal impact evaluations are being made.

If the *F*-test indicates that the null hypothesis is almost certainly wrong—that is, in our simple case, there is a difference between the zones that cannot be explained by chance variations—there remains the important question as to which zone or zones are the “odd man out.” Does one particular zone stand out as having different farmer behavior from the others, which are similar, or do all the zones have idiosyncratic farmer performance? A simple method of completing the analysis is to use a technique known as Tukey’s significant gap. We assemble the zonal means in ascending order and calculate the gap between successive zones. An example is shown in table 19.

Let us assume that an analysis of variance has been carried out. The value of *F* was significant at the 95 percent point confidence level. The computer printout also shows that the “within-zone mean square” as the denominator in the *F* calculation is 3,823. A significant gap is given by

$$G = \sqrt{2} \times t_{0.05} \times \sqrt{\frac{\text{within-zone mean square}}{n}}$$

where $t_{0.05}$ is the 95 percent point of the *t*-distribution, which as stated earlier is approximately 2; and *n* is the number of observations in each zone, which in this case is 20.

Therefore, for our example,

$$G = \sqrt{2} \times 2 \times \sqrt{\frac{3,823}{20}} = 39$$

Table 19. Calculation of Gap between Zones

Zone	Mean fertilizer application rate (kg./ha.)	Gap
3	90	20
1	110	
4	120	
2	160	

So zone 2 is the one that differs significantly from the others with a gap of 40.

There are two other important limitations in interpreting an F -test from an analysis of variance. First, the F -value does not in itself denote the proportion of variation that is explained by the variable being tested. If two variables both result in statistically significant F -values, the fact that one may be larger than the other is of no practical value. Second, if the data set is large, an analysis of variance of almost any possible explanatory variable is likely to produce a significant result. The chance that the zones, let us say, are absolutely equal in their farmer behavior is small. With sufficient data we are almost bound to conclude that there are differences, but whether these differences are of any practical importance from the point of view of tactical or strategic project implementation is a very different matter.

We therefore conclude that if an analyst is not well versed in the technique he would be well advised to steer clear of it, but even when carried out, the results should be viewed with caution. Most monitoring and evaluation reports will lose little if the analysis of variance is omitted. What must not be omitted is sensible presentation of the basic tabulations and cross-tabulations. Most of us would have singled out zone 2, for example, without the aids described. We say more on this topic in the next chapter.

Regression and Correlation

No weapons in the evaluator's armory are more commonly relied on than the fitting of linear regressions and the calculation of correlation coefficients. Econometricians are even more dependent on these measures of association. This section introduces these methods; the next section warns of their dangers, for they are commonly misused. We will develop the regression method using the classical least-squares approach, but the reader is referred to chapter 8 for alternative methods which are quicker and in certain circumstances more robust.

A linear relation between two variables y and x can be expressed in the form $y = a + bx$, where a is the value of y when $x = 0$, and b measures the slope of the line that represents the relation. In any practical project evaluation situation, the relation will not be perfect—deviations will be observed which are caused either by random variations between the observed cases or errors in the observations themselves. So we introduce the residual term (ϵ_i) and write the expression for any individual observation of the two variables as $y_i = a + bx_i + \epsilon_i$ where ϵ_i has a mean of zero.

The least-squares method (which involves differential calculus) minimizes the sum of the squares of the deviations of individual observations from the fitted line. It results in the following formula for b , and then a :

$$b = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$a = \bar{y} - b\bar{x}$$

Figure 11 shows the plot of the proportion of contact farmers adopting a new crop variety by year of project implementation. Table 20 shows the calculation of b and a if done without the aid of a calculator.

Such a relation is used to estimate what the value of y would be for a given value of x if the random residual did not intervene. Or in a case such as that shown in figure 11, the value of y for a value of t (that is, time) has not yet occurred: that is, we use the relationship to forecast the value of y in future seasons or years. Its usefulness for these purposes depends on how close the actual values are to the fitted regression—how large are the values of ϵ_i . The measure of the goodness of fit is known as the correlation coefficient between x and y and is denoted by r where

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

The extra manual calculation needed for this is shown in the last column of table 20.

Figure 11. Regression of Proportion Adopting by Year

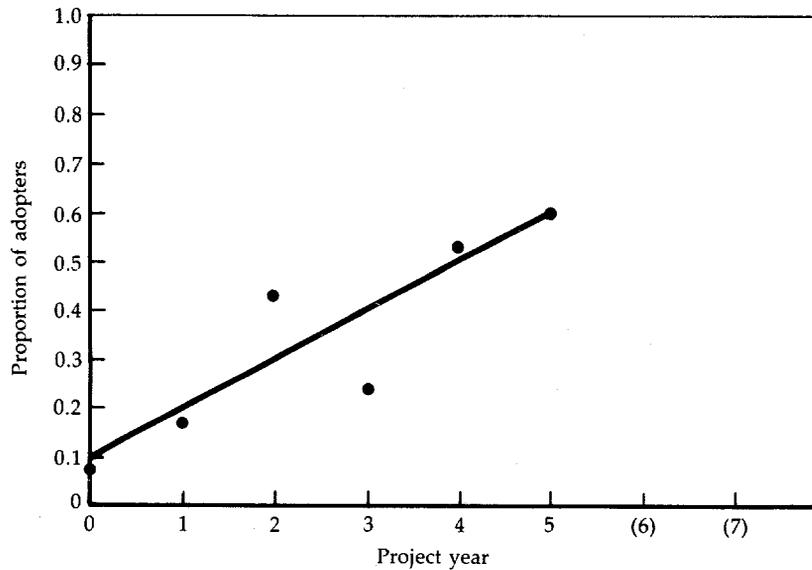


Table 20. Method of Manual Calculation of b and a

Year (x)	Proportion of adopters (y)	$x - \bar{x}$	$y - \bar{y}$	$(x - \bar{x})(y - \bar{y})$	$(x - \bar{x})^2$	$(y - \bar{y})^2$
1	0.17	-2	-0.22	+0.44	4	0.0484
2	0.43	-1	+0.04	-0.04	1	0.0016
3	0.23	0	-0.16	0	0	0.0256
4	0.52	+1	+0.13	+0.13	1	0.0169
5	0.60	+2	+0.21	+0.45	4	0.0441
Total				0.95	10	0.1366

Note: $b = 0.95/10 = 0.095$; $a = 0.39 - 0.095 \times 3 = 0.105$. The line fitted through these data is then expressed by the formula $y = 0.105 + 0.095x$.

The value of r will lie in the range of $-1 \leq r \leq +1$. Negative values of r mean that the regression line is downward sloping, that is, y decreases as x increases. The closer r is to either -1 or $+1$, the better the regression line explains the location of the observations. If r is close to zero, the estimated association between x and y is a very poor predictor of the value of y for a given x . For the data shown in figure 11, $r = 0.81$. The linear relationship therefore would seem to provide a good fit to the observed points.

Dealing with only two variables is an oversimplification of the situation encountered in evaluation work. No one single variable x may explain the behavior of y , but there may be a series of variables, x_1, x_2, \dots, x_n which together jointly determine changes in y . We generalize our model to the multivariate linear regression case, $y = a + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$. The calculations of a and the set of b_i s, ($i = 1, \dots, n$) are more complicated than in the bivariate case; they require the aid of computing facilities.

Of course there may be a true relation between y and x_i ($i = 1, \dots, n$) but not a linear one. The theory of statistics can accommodate this with curvilinear regression, but the use of this lies outside the needs of most evaluation work.

Returning to the coefficient of correlation, r , it can be shown that r^2 is an estimate of the proportion of the total variation in the values of y that can be explained by the fitted relation with x . In the case of the example in figure 11, $r^2 = 0.66$, so we may say that approximately two-thirds of the variation in the number of farmers adopting the new variety is explained by the time factor—the number of years since project inception.

The F -test, using the analysis of variance, explained earlier, is used to determine the significance of the regression relationship; but as stated earlier, if the data set is large the likelihood is that significance will be established. This does not mean that the relationship is always of opera-

tional value, particularly when r^2 is small. We also advocate caution when interpreting r^2 values that are very high, say more than 0.75. Such strong relations explaining most of the variation in the data rarely occur in real project circumstances (as opposed to experimental trials). In such circumstances, the evaluator is well advised to investigate the validity of the data.

We are using r (or r^2) as a measure of the relationship between x and y . If $r = 0$, we are tempted to say that there is no relationship and that x and y therefore are statistically independent. If variables are normally distributed, it is true that they are independent if and only if they are uncorrelated. But if variables are not normally distributed, r cannot be used to determine whether they are independent or related. In other words, r is only useful as an indicator when the variables are distributed normally. Because of this it is common practice in econometrics and regression analyses for evaluations to transform skewed data so as to introduce normality, or an approximation of it. Taking the square root of one or both variables or alternatively the logarithmic values are the commonest transformations used. Even if variables are normally distributed, it does not follow that a high r^2 proves that there is a relation between them that has practical meaning for project implementation. This brings us to the warning section.

Cautionary Comments about Correlation and Regression

The data analyst needs to formulate a hypothesis regarding a relation to be examined before using correlation and regression methods. There is a time-honored methodological maxim: "Correlation does not prove causation." Correlation of an effect with an identified stimulus is only useful evidence if the other causal factors have been factored out. Production may have increased as fertilizer sales rose, but the real stimulus to production gains may have been changes in market prices. The correlation of two variables observed over a number of time points is particularly likely to lead to mistaken inferences. A high correlation coefficient may be caused merely by the separate high correlations of each variable with time. There are famous examples of nonsensical correlations in excess of 0.9 between, for example, salaries of clergy and sales of alcoholic beverages when in fact both rise fairly steadily over an extended time. We have said in chapter 8 that the first differences (period-to-period changes) should be correlated—if the value of r for differences is high, the likelihood of a real relation between the variables, independent of time, is increased.

Another common error in interpreting correlation coefficients between variables stems from the common situation in which a significant corre-

lation measured over aggregates of persons does not appear when the correlation is calculated over individuals. There may be a correlation at an aggregated level (say, by district) between educational attainment and business activity, but when the data are examined at the individual level there may be little correlation between education and commercial or industrial activity. Districts with booming business tend to contain high numbers of better educated people, perhaps because a high standard of living facilities tends to accompany economically booming areas and better educated people are attracted to areas of high living standards. Indeed, a general problem with correlation coefficients is that their size can be increased by increasing the level of aggregation in the data used to calculate them.

Nor is it sufficient to show that the value of r^2 is "statistically significant." Many such statements are of little interest or importance for policy purposes.

Perhaps the most serious problem in using regression models for data sets collected in a real project environment is the effect of measurement errors. We have commented already that the use of the least-squares method, which underpins the calculation of b , the regression coefficient, requires that the independent variable x be observed or measured without error. If this condition does not hold, the reliability of the estimate of the regression coefficient is seriously affected. Quick methods of fitting the regression line were discussed in chapter 8; these may give more valid estimates when errors affect both the dependent and independent variables.

These errors are not sampling errors but rather the difference between the true value for a particular case and the figure that is actually recorded and used in the analysis. Consider the situation in which the value recorded for the i^{th} case, X_i , does not necessarily equal the true value of X_i , the error being equal to w_i . Thus $x_i = X_i + w_i$.

The w_i s can be positive or negative, and of course can be equal to zero. Most commonly, errors are introduced during the collection of the data in the field, but they also arise during the editing and coding of the data. It is a rare data set dealing with the evaluation of agriculture projects that does not suffer from a set of w_i s of considerable size. In a simple one-independent-variable regression model, the "true" value of the coefficient b is related to the calculated value \hat{b} by the equation $b = \hat{b}/R$, where R is the reliability of the measurement of the independent variable. This result is sometimes referred to as the attenuation effect:

$$R = \frac{\text{variance of } X}{\text{variance of } X + \text{variance of } w}$$

The variance of w may be expressed as a proportion of the variance of X : $\text{var } w = k \text{ var } X$. Then it follows that $R = 1/(1+k)$ and $1/R = 1+k$.

Thus if $k = 0.1$, the calculated value \hat{b} should be multiplied by 1.1 to correct for the attenuation effect. But in many cases of poorly observed values or estimates given by respondents on the basis of long recall periods, k may be 0.3 or more. With regard to multiple regression, where we are dealing with a set of independent variables $x_i, i = 1, \dots, n$, the issue of errors in these variables is more complicated.

A common ambition in evaluation is to analyze an impact variable such as income or yield by regressing it on a set of projected-related and exogenous variables and choosing a minimum set of these variables that appear to be the ones that most influence the impact variable. Choosing this subset by the use of stepwise regression techniques is a popular type of analysis, for computer software packages provide the means of handling the daunting volume of calculations. In this volume and its companion we have argued against such ambitious evaluations on other grounds. A further reason is that many of the analyses of this type are invalid. It is only recently that competent statisticians have addressed the growing popularity of such analyses and the fundamental problems they present. The following quotations are from one discussion of this issue³: "Simple selection methods fail to deliver, the usual significance tests are misleading, estimated regression coefficients are biased." "Variable selection methods using t or F criteria are really aiming to eliminate x -variables whose true coefficients are exactly zero. The implausibility of the occurrence of such variables in practical situations is another argument against the appropriateness of these methods."

An exposition of the theoretical problems and practical dangers in using the stepwise multiple regression to which these speakers allude would take us way beyond the scope of this chapter. But such problems exist, which emphasizes again that our recommendation for simple analysis may not be second best but rather all that our error-prone data can support.

3. A. J. Miller, "Selection of Subsets of Regression Variables." *Journal of the Royal Statistical Society*, series A, vol. 147, no. 3: 410-15.

10 | *Presenting Data to the User*

IF THE PROCEDURES DESCRIBED in this book, particularly in chapters 8 and 9, have been followed, the required amount of data will have been collected, sorted, collated, and marshaled into a logical structure; patterns will have been explored; and hypotheses will have been tested. Now the crucial step is to present the data and accompanying text crisply, succinctly, and on time. Chapter 6 of the companion volume discusses communicating information, including report writing. Within that overall framework, this chapter discusses the issues that arise in presenting the data.

The presentation of data to managers at various levels is facilitated if the following principles are observed:

- Levels of detail and disaggregation will vary so as to be appropriate for a particular type of reader.
- Definitions of variables, table headings, and layouts will be made clear to the reader, who will not always have either a mathematical background or technical knowledge of the topics discussed.
- References to the technical terms used in statistical analysis will be explained for the nonprofessional reader.
- Text accompanying tables will summarize the main highlights revealed by the tables, indicating the conclusions that may be drawn.
- Graphics and other diagrams will be used to focus the reader's interest and aid his understanding.

Those undertaking to transmit data should be willing to communicate provisional and approximate findings if they are significant. Unfortunately, those who are fearful of making a mistake tend to err on the side of caution. Careful checking of surprising facts, deliberate reviewing of the quality of data, and not jumping to premature conclusions are all valuable attributes; but some willingness to make the best judgment possible on the basis of whatever evidence is currently available is also necessary.

Previous chapters have introduced both exploratory and formal statistical analyses of data. The following sections thus discuss only the tabular and graphic presentation of data and analytical findings.

Tables and figures should allow the reader to absorb readily the composition of the data set and appreciate without further analysis the most obvious patterns and relationships.¹ Some techniques to help achieve this include the following:

- Use clear and unambiguous class intervals in frequency distributions.
- Truncate the number of digits shown.
- Use clear, self-explanatory column and row headings.
- Liberally use differential spacing to highlight comparisons.
- Order columns and rows sensibly.
- Transform the data into percentages and indexes as appropriate—without forgetting to apply techniques already mentioned to the transformed data.
- Use distributional parameters—averages, standard deviations, and so on, to summarize the arrays of data.
- Compare the data with data from earlier years or other geographical regions.
- Use simple, uncluttered pie charts, histograms, and graphs to highlight features of interest.

Frequency Distributions

When a large data set is being summarized into a grouped frequency distribution, the choice of class intervals is the crucial decision. Consider the equal class-width distribution of farm sizes in table 21. It might be more informative to split the first two classes into one-hectare widths because 60 percent of the total data set is in these classes. Also, in summarizing the data, the upper classes could be merged into one group of 14.1–20 hectares.

There is a danger, however, in the excessive use of unequal class intervals. The reader who sees a distribution presented in this form is first inclined to see it as a set of discrete and equal steps; only close observation will reveal the varying widths of the classes. Our recommendation is to maintain equal class intervals through most of the distribution and use wider intervals if required for the last one or two classes.

1. For further treatment, see A. S. C. Ehrenberg, *Data Reduction* (London: Wiley, 1978).

Table 21. Frequency Distribution of Farm Sizes

<i>Farm size (ha.)</i>	<i>Number of farms</i>
0-2	350
2.1-4	250
4.1-6	150
6.1-8	100
8.1-10	70
10.1-12	30
12.1-14	25
14.1-16	15
16.1-18	5
18.1-20	5
<i>Total</i>	1,000

Note also how the farm size class intervals are referenced, 0-2, 2.1-4, and so on, with the definition of the unit, hectares (ha.), included in the row heading. This implies that any farm of more than 2 ha. but not larger than 4 ha. would be in the second class, but strictly it could be argued that a 2.02 ha. farm would be included in the 0-2 ha. class because its area would round to one decimal place as 2.0 ha. Unless one is very pedantic and shows the class boundaries defined to the same precision as the original recording of the values, there is always some small level of ambiguity. But to avoid giving the table a messy appearance, this is usually acceptable. It is not uncommon to find the classes shown as, say, 0-2, 2-4, 4-6—which leaves uncertain where a farm of precisely 2, 4, or 6 ha. is placed. With continuous variables, for which very few of the values will be precisely measured on the class interval, this may be acceptable. But when discrete variables such as household size are being shown, this ambiguity must be avoided.

How many classes to show is a matter of choice. Obviously, the point of a grouped distribution is to summarize the data so that the eye of the reader assimilates the essential shape—where the mode is and how quickly it tails off. Ten classes, as in the above example, seems to us to be a reasonable maximum to aim at.

Truncation of Digits and Summary Distributions

Most tables presented to managers fail to communicate because numbers are shown to an excessive number of digits. Drastic surgery is often needed to round the values by removing decimal places or units (tens and even hundreds). Little is lost (unless, exceptionally, the precise value of a carefully observed characteristic is required), implications of spurious ac-

Table 22. Maize and Bean Yields by Farm Size and Project Participation

Participant	Farm size (ha.)	Yield (kg./ha.)	
		Maize	Beans
Yes	2.43	2,762	941
Yes	6.12	2,248	1,060
Yes	5.43	1,710	812
No	2.17	2,361	743
No	1.64	946	671
Yes	7.43	2,432	1,165
No	4.80	2,170	846
Yes	3.71	3,432	—
Yes	2.55	1,487	504
No	1.75	1,700	764
No	1.04	619	740
No	3.27	1,860	910
Yes	12.14	1,640	831
No	4.61	1,940	663
No	0.93	835	—
Yes	2.10	3,471	—
Yes	5.12	2,743	1,230
No	1.79	1,876	910
No	1.29	1,149	—
Yes	2.91	4,140	1,423

— No cases.

Table 23. Maize and Bean Yields by Project Participation

Participants			Nonparticipants		
Farm size (ha.)	Yield ('000 kg./ha.)		Farm size (ha.)	Yield ('000 kg./ha.)	
	Maize	Beans		Maize	Beans
2	2.8	0.9	2	2.4	0.7
6	2.2	1.1	2	0.9	0.7
5	1.7	0.8	5	2.2	0.8
7	2.4	1.2	2	1.7	0.8
4	3.4	—	1	0.6	0.7
3	1.5	0.5	3	1.9	0.9
12	1.6	0.8	5	1.9	0.7
2	3.5	—	1	0.8	—
5	2.7	1.2	2	1.9	0.9
3	4.1	1.4	1	1.1	—

— Crop not grown.

Table 24. Maize and Bean Yields by Farm Size

<i>Farm size (ha.)</i>	<i>Participants</i>		<i>Farm size (ha.)</i>	<i>Nonparticipants</i>	
	<i>Yield ('000 kg./ha.)</i>			<i>Yield ('000 kg./ha.)</i>	
	<i>Maize</i>	<i>Beans</i>		<i>Maize</i>	<i>Beans</i>
2	3.5	0.5	1	0.8	—
2	2.8	0.9	1	0.6	0.7
3	1.5	—	1	1.1	—
3	4.1	1.4	2	0.9	0.7
4	3.4	—	2	1.7	0.8
5	1.7	0.8	2	1.9	0.9
5	2.7	1.2	2	2.4	0.7
6	2.2	1.1	3	1.9	0.9
7	2.4	1.2	5	1.9	0.7
12	1.6	0.8	5	2.2	0.8

— No cases.

Table 25. Maize and Bean Yields Grouped by Farm Size

<i>Farm size (ha.)</i>	<i>Maize yields ('000 kg./ha.)</i>		<i>Bean yields ('000 kg./ha.)</i>	
	<i>Participants</i>	<i>Nonparticipants</i>	<i>Participants</i>	<i>Nonparticipants</i>
Less than 2	—	1.2 (6)	—	0.8 (4)
2-3	3.1 (5)	2.1 (2)	1.0 (3)	0.8 (2)
4-5	2.2 (2)	2.1 (2)	1.0 (2)	0.8 (2)
6+	2.1 (3)	—	1.0 (3)	—
<i>Total</i>	2.6 (10)	1.5 (10)	1.0 (8)	0.8 (8)

— No cases.

Note: Numbers in parentheses denote the number of cases on which the yield is based.

curacy are avoided, and much is gained in reader understanding. Moreover, having looked at the data in truncated form we may find that the main message can be conveyed with a very abbreviated tabular format. Table 22 illustrates this point. For ease of presentation we have kept the number of values low, but the principles apply equally to larger data sets.

The various presentational possibilities illustrated below are not intended to imply that all versions are meaningful or necessary to include in a formal presentation. The examples show the range of presentational possibilities, from which the most revealing would be chosen.

First, for presentational purposes, let us (in table 23) truncate and round the data to the nearest hectare and assemble the project participants in one column and the nonparticipants in another.

Now let us (in table 24) order the data by farm size.

A picture now starts to emerge. Project participants tend to have larger farms and obtain better yields. Let us now (in table 25) form a two-way table by grouping farm sizes, calculating the mean yields, and bringing the comparative yields together. (For calculation of the means, we use the untruncated data and note that four farmers do not grow beans.)

Differences in yields between project participants and nonparticipants seem to be caused by factors other than farm size. Although the participants tend to have larger farms, the maize yields achieved on the larger farms are no better than others, and bean yields do not vary with farm size.

Ignoring farm size, let us examine the grouped frequency distribution of maize and bean yields:

	<i>Maize yields</i> (<i>'000 kg./ha.</i>)			
<i>Farmers</i>	<i>Less than 1</i>	<i>1-2</i>	<i>2-3</i>	<i>3+</i>
Participants	—	3	4	3
Nonparticipants	3	5	2	—

	<i>Bean yields</i> (<i>'000 kg./ha.</i>)			
<i>Farmers</i>	<i>Less than 0.8</i>	<i>0.8-1</i>	<i>1.0-1.2</i>	<i>1.2+</i>
Participants	1	3	2	2
Nonparticipants	5	3	—	—

These simple tables convey much of what the data have to reveal. If anything more is needed, it will be a quantification of the obvious distributional differences:

	<i>Maize yields</i>		<i>Bean yields</i>	
<i>Farmers</i>	<i>Mean</i> (<i>'000 kg./ha.</i>)	<i>Standard</i> <i>deviation</i>	<i>Mean</i> (<i>'000 kg./ha.</i>)	<i>Standard</i> <i>deviation</i>
Participants	2.6	0.9	1.0	0.3
Nonparticipants	1.5	0.6	0.8	0.1

We could now conduct *t*-tests to confirm the significance of the difference between participants and nonparticipants, but the visual impact of the first two tables in the last paragraph has achieved the purpose. Table 25 may be preferred to the third table in the last paragraph in this example, for the interaction between farm size and participation should not be suppressed.

Ruthless truncation of decimal places and insignificant figures plus the appropriate use of grouped frequency distributions can reveal most of what the analyst wishes to convey without any formal statistical analysis whatsoever. Nor is much lost by such truncation. Even if the recorded values are precise, rounding does not reduce the appreciation of variability in the data unless the data are very homogeneous. For many data sets collected for monitoring and evaluation purposes, the observational errors are at least as great as the level of truncation. If a yield is measured as 3,642.51 kilograms per hectare with the use of a crop cutting sample technique, to quote the figure in this form is to give a completely erroneous impression of accuracy. It is better to show it in a table as "3.6 thousand kilograms," which is easier for the eye to scan and avoids giving the wrong impression of a high level of accuracy in the measurement.

Presentation of Cross-Tabulations

The two-way table is the commonest form in which to present data. The analytical aspects of cross-tabulations were discussed in chapter 9. Here we consider presentation.

Two types of tables are involved: one in which the entries in the cells of the tables are counts, the frequency with which a case belonging to that cell occurred; the second in which the cell entry is a value such as a mean, percentage, or total—for example, mean yield, percentage adopting, total area planted.

Both types have some principles in common. The column and row headings should be clear and unambiguous. The spacing, particularly between columns, should help the figures stand out as discrete entries rather than as a blur of continuous digits with but the narrowest of spaces between the end of one entry and the start of the next. Truncation of digits is a major help in achieving this spacing. Particularly important columns or rows can be highlighted by the use of differential spacing, and, if good quality printing is available, by the use of darker type.

If one of the dimensions of the tabulation is geographical—for example, project zones or districts—its ordering should be considered carefully. Commonly, geographical characteristics are listed in their actual physical sequence, say from north to south, or in the sequence customarily used in national statistical digests; but in some cases it may be useful to list them in the order of their importance in relation to the variable shown

Table 26. Milk Production by Province

<i>Province</i>	<i>Total production (millions of liters)</i>	<i>Number of holdings with cows (‘000)</i>	<i>Number of cows in milk (‘000)</i>	<i>Production per cow (liters)</i>	<i>Percentage of production sold</i>
Central	310	210	333	932	46
Eastern	172	193	446	386	29
Rift Valley	153	77	290	529	36
Nyanza	138	194	525	263	38
Western	48	118	168	283	31
Coast	18	20	85	213	62
<i>Total</i>	839	812	1,847	455	38

Note: Excludes pastoral and larger farm areas.

Source: Central Bureau of Statistics, Kenya, *Report of Integrated Rural Survey 1974-75* (Nairobi, 1975).

Table 27. Number of Fertilizer Users

Farmers	Zone						Total
	A		B		C		
Participants	276	36	180	23	314	41	770
	73	(27)	80	(18)	78	(31)	76
Nonparticipants	102	43	46	19	90	38	238
	27	(10)	20	(5)	22	(9)	24
<i>Total</i>	378	38	226	22	404	40	1,008

in the table. For example, if the table is showing crop areas by region, it may help the reader to list the regions in descending order of total area under cultivation. Table 26 shows a cross-tabulation with these desirable features.

If the cell entries are counts rather than values, the opportunity exists to show both the numbers themselves and the percentage distribution by row, column, and by row and column. Some statistical packages show all these combinations in one cell. Table 27 is a simple hypothetical example. The large numbers are the actual counts, the figures to the right of the count are the row percentage distributions, the figures under the counts are the column percentage distributions, and the figures in parentheses express the overall row-by-column percentage distribution. Such a table may be useable by the analyst, but it will totally confuse most readers. Tables should not be as cluttered as this one. The count and either the row or column percentage should be the maximum number of entries in a cell. The choice of the percentage depends on the message to be conveyed (see the next section of this chapter).

When the cell entries are values, the most important point to remember in producing a final version of the table is that the marginals (row and column entries) are not subtotals as in the type just discussed. Therefore, row and column percentages are not possible; this point, obvious though it is, is sometimes overlooked.

The presenters' main responsibility in a cross-tabulation of this type is not to stretch the data beyond reasonable limits. Cells with values that are based on one, two, or three cases may be dangerously misleading. If the number of cases in certain cells is limited, it is advisable to indicate in footnotes the existence of such cells or to suppress the values by inserting a marker denoting "insufficient number of cases." An alternative is to indicate in brackets the number of cases in each cell—if this can be done within the rules for clarity of presentation already discussed.

Consider table 28, an actual table based on a total sample of 71 cases.

Two of the cells show yields based on one observation only. Others are based on two or three observations. Because of the insufficient number

Table 28. Correlation between Fertilizer Use and Rice Yields

<i>Farm size and type of rice</i>	<i>Average yield (ton/ha.) for Different Application Rates</i>				
	<i>Less than 126 kg./ha.</i>	<i>126-188 kg./ha.</i>	<i>189-250 kg./ha.</i>	<i>251-312 kg./ha.</i>	<i>more than 312 kg./ha.</i>
Small	2.4	4.9	4.3	3.8	4.5
Medium	3.2	4.4	4.1	4.1	4.6
Large	4.7 ^a	4.3	3.8	—	—
Transplanted	4.7 ^a	5.1	4.1	3.8	4.1
Broadcasted	3.0	4.3	4.0	3.9	4.8
Total	3.1	4.4	4.0	4.0	4.6
Average					

— No cases.

a. Information obtained from one farmer only.

of cases, no uniform trends exist along the rows and the impression is created that yields are semi-independent of fertilizer application rates. This table conveys no meaningful message because the data have been overstretched; the table should have been suppressed or compressed.

Notice also the ambiguity in the row heading—the rows seem to be nonexclusive and based first on size of farm and then on sowing practice, but this is not very clearly stated. Also, the heading “Total Average” is scarcely to be recommended, particularly when the words appear underneath each other giving the impression that the entries are “Totals” and that the “averages” are missing. “Weighted mean” or “Overall mean” would be preferable.

Presenting Percentages

Percentages and other indexes can either be very helpful in presenting data clearly or very misleading, depending on the competence (and honesty) of the presenter. Consider the data in this table giving the number of farmers purchasing fertilizer in each of three districts in each of three years:

District	Total farmers	Farmers purchasing fertilizer		
		Year 1	Year 2	Year 3
A	10,000	3,612	4,170	4,670
B	14,000	765	1,241	2,073
C	38,000	21,036	20,117	19,416

We can present these data in terms of the percentage of farms in a district purchasing fertilizer in each year (noting that each percentage is independent of the others—they do not sum to 100 in any direction):

District	Percentage of farmers purchasing fertilizer		
	Year 1	Year 2	Year 3
A	36	42	47
B	5	9	15
C	55	53	51

Or we can present the percentage distribution by district of the farmers buying fertilizer in a year:

District	Percentage of fertilizer purchasers		
	Year 1	Year 2	Year 3
A	14	16	18
B	3	5	8
C	83	79	74
Total	100	100	100

Or we can show the index of the number purchasing fertilizer in each district, taking the number in year 1 as the base (100):

<i>District</i>	<i>Index of fertilizer purchasers</i>		
	<i>Year 1</i>	<i>Year 2</i>	<i>Year 3</i>
A	100	115	129
B	100	162	271
C	100	96	92

The first table in the last paragraph shows that fertilizer is most popular (or more easily available) in districts A and C; it also shows the time trend, but it disguises the dominance of district C in terms of numbers of purchasers. The second table shows this dominance of district C but disguises the trends in numbers of purchasers over time. The third table highlights the time shifts most clearly, bringing out the rapid growth in district B and the decline in district C while losing the relative importance in absolute numbers or relative incidence of fertilizer purchasers in the various districts. Each such presentation has a role to play, but each presents a partial picture, which in isolation can even mislead.

In a two-way table the choice of the direction in which to calculate the percentage distribution may be important. Consider an example of farmers in three locations classified by the number of parcels of land making up the holding:

<i>Number of parcels</i>	<i>Location</i>			<i>Total</i>
	<i>A</i>	<i>B</i>	<i>C</i>	
1	100	20	16	136
2	60	60	24	144
3	40	40	40	120
<i>Total</i>	200	120	80	400

Little may be detected from a casual examination of this, except that there are more farmers in area A. Consider, however, the following table, which expresses the same data in percentage form for each area:

<i>Number of parcels</i>	<i>Location</i>		
	<i>A</i>	<i>B</i>	<i>C</i>
1	50	17	20
2	30	50	30
3	20	33	50
<i>Total</i>	100	100	100

This table clearly reveals that farmers in the three areas have very different distributions in terms of fragmentation of holdings, with the modal number being 1, 2, and 3, respectively. Conversely, to run the percentages across the rows would reveal nothing, except possibly the larger size of area A. Simplistic though these examples are, there have been many real cases where a wrong choice of the direction for calculating percentages masked the story that the figures had to tell.

Graphs and Figures

Chapter 8 discussed the use of graphs by the analyst as an aid to exploring and understanding data. We now turn to their use in communicating findings. The purpose of a pictorial presentation is to dramatize a point without deliberately deceiving the reader's eye. Too often either the potential of a pictorial impact is missed because of the selection of an inappropriate graph or excessive cluttering of its contents, or, by selecting false origins, as explained below, or other misrepresentation—such as undue expansion of the vertical scale—it can become exaggerated.

In general, graphs and figures are useful in conveying a trend or comparison of relative sizes in various categories. They are poor in conveying actual numbers even if the numbers are entered on the graph.

Figures 12, 13, and 14 are examples of a graph, a bar chart, and a pie chart. They give the eye an important message that the presenter of the data wishes to convey. They are clear and uncluttered, and they dramatize the point that is being emphasized without undue distortion.

But consider figures 15 and 16. Figure 15 is an example of excessive cluttering of a graph resulting in the eye obtaining an image of confusion and woven threads rather than the significant differences between regions. And figure 16 is an example of the commonest technique for exaggerating a message—namely, truncating the lower portion of a graphic scale. The decline shown is approximately 64 points from a base of 2,156—less than 3 percent. This example of a decline in a stock market index was indeed unusual, but because the vertical scale has a starting point of 2,080 the decline takes on the appearance of total collapse. If the graph is drawn with zero as the baseline, however, any effect is nullified because much of the surface area of the graph will be occupied by that part of the scale that the index had never reached in modern times; the message of an unusual day thus would be totally masked. Where then should the line literally be drawn? Many would argue that no nonzero base can be justified, so that a graph is inappropriate in such cases. It is a matter of judgment in each case. Our view, in the present example, would be to start the scale at the lowest level of the index in, say, the previous year. This would put the one-day decline in better perspective.

Attribution of Accuracy and Significance

The presenter of data should discuss the likely accuracy of the data and indicate which findings are significant in view of the margins of error involved. The analysis of administrative records may present little difficulty if based on an examination of the total file, which supposedly includes all relevant cases. The quality of the record keeping may, however, be variable, and such variations should be noted.

Figure 12. Sales of Maize by Credit Recipients

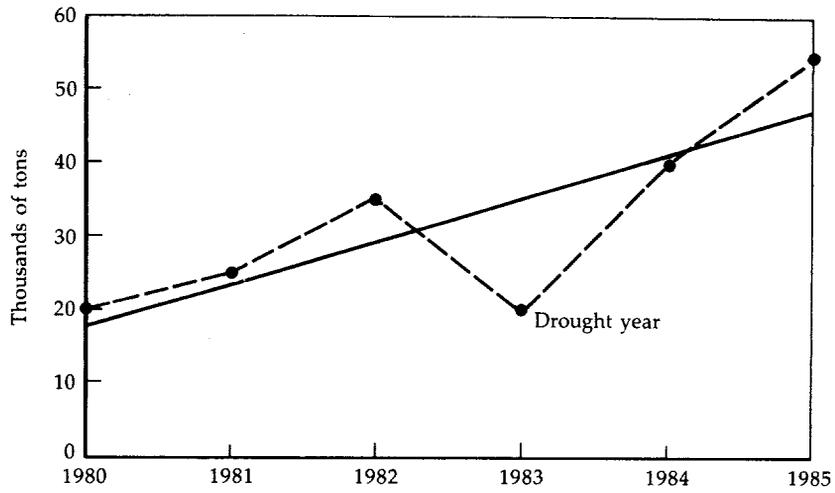
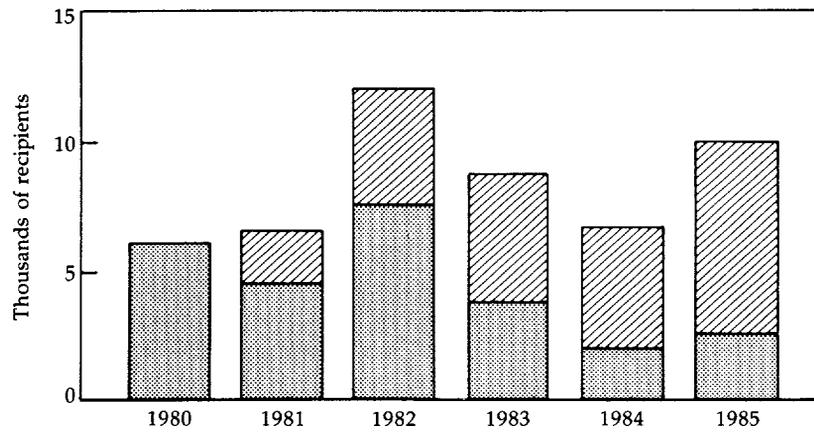
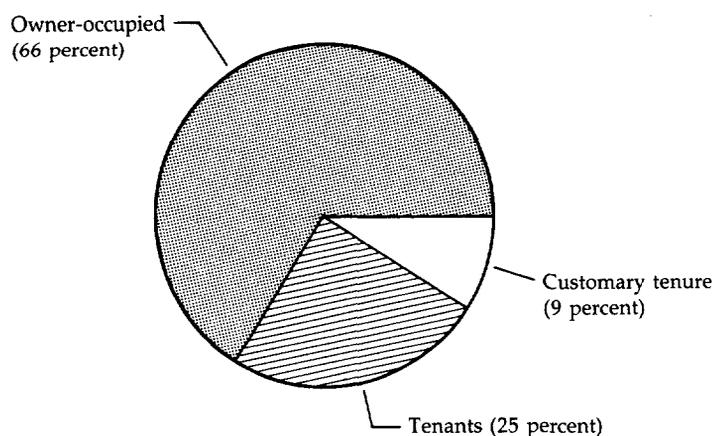


Figure 13. Credit Recipients by Year



Key: [hatched box] repeat borrowers [stippled box] first-time borrowers

Figure 14. Credit Recipients by Land Tenure



Intensive, small-scale studies present the reporter with a difficult decision regarding the general validity of the findings. It will not usually be possible to make inferences from a limited number of cases (perhaps purposively selected) to the population at large. This is certainly true for numerical estimates of totals, ratios, and so on. A case study, however, aims to analyze the internal relationships of the variables that cause a particular phenomenon. If a consistent pattern emerges, plausible inferences may be made about the phenomenon's mechanism, which may be likely to repeat under similar circumstances. As long as a suitable cautionary note is struck, general inferences of causality and relationships may be made on the basis of plausibility. For example, case studies may examine in depth a few farmers who adopted or rejected a certain project recommendation. Say that it is found that the reluctance to adopt the new farming practice was quite logically caused by the perception that labor will be constrained when the new practice makes heavy demands. Assume also that the adopters have more family labor available than the nonadopters and did not perceive the same constraint. The conclusion that the choice of adoption or nonadoption is affected by the availability of family labor may be a plausible inference for the general population in the area, even though the numbers affected by the limited availability of labor cannot be calculated from the study.

The communication of such findings requires skill in interpretation and presentation. The temptation to claim too much must be resisted, particularly when the pattern is inconsistent and blurred by a range of extraneous factors. It may be a worthwhile conclusion that no discernible

Figure 15. Percentage of Agriculture Lending by Region and Year

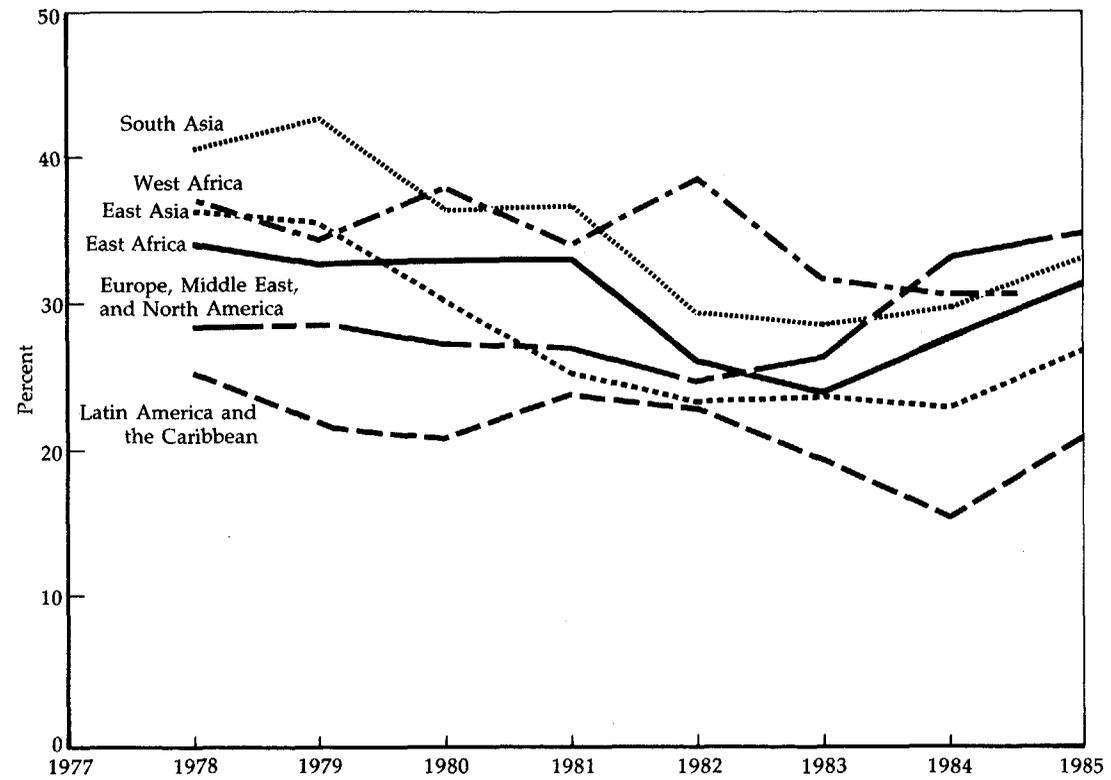
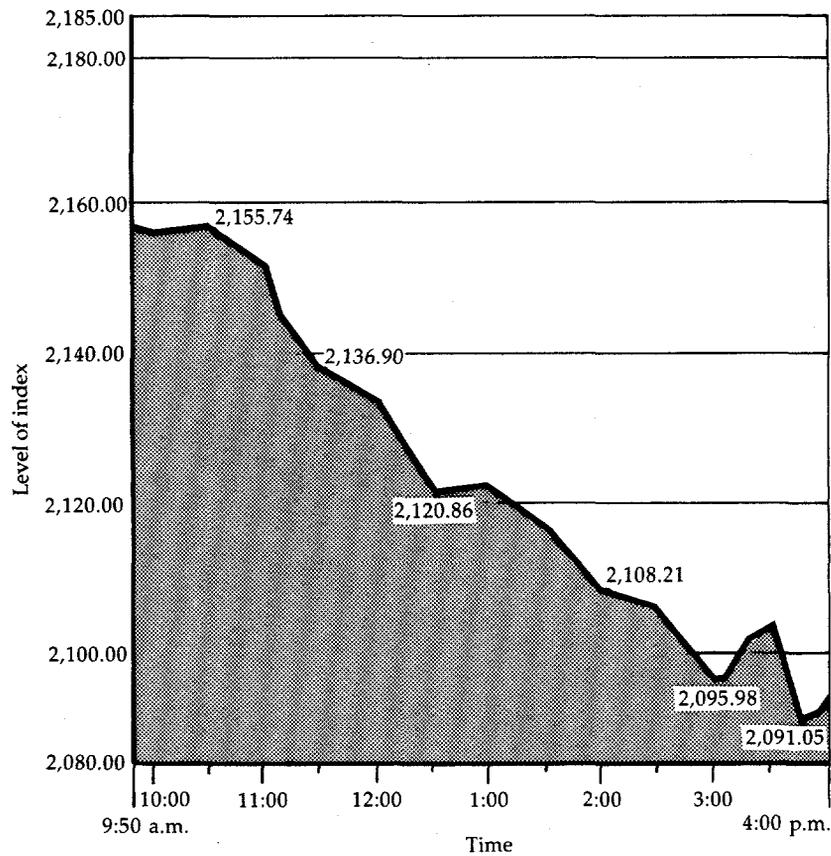


Figure 16. Decline of a Stock Market



pattern exists—then at least one set of hypotheses can be rejected, which will help managers to avoid concentrating on the removal of an irrelevant constraint.

Results from properly designed and executed sample surveys enable numerical inferences to be made to the population at large with calculable margins of sampling error. The procedures for doing so were introduced in chapter 6, although from the perspective of calculating the required sample size for a given sampling error. The formulas for calculating the margin of error for any sample design can be found in standard texts on sample surveys.²

Once the sampling error is calculated, the question arises of how to determine which estimates are significantly different—either from one an-

2. See, for example L. Kish, *Survey Sampling* (New York: Wiley, 1965).

other or from a standard of comparison— so that conclusions may be drawn with reasonable confidence. The techniques have been introduced in chapters 8 and 9, and further elaboration can be found in standard texts. What we note here, as in the companion volume, is the danger that the reporter may set too demanding a standard before describing a result as significant. The use of the classic 95 percent confidence level implies that any statement of a significant change or difference has a very high chance (19 to 1) of being correct. Without a recommendation emerging from the survey, an existing procedure may be maintained when there is, let us say, a 3 to 1 chance that a change would be beneficial. Or if a decision must be made, the manager may be forced to rely on subjective judgments and be unaware that the data showed strong, but less than certain, support for one particular course.

The relative costs of making an unneeded change compared with those of leaving a procedure in place that required changing are the determinants in deciding the level of confidence to be attached to testing for significance. In one case, the cost of making a change may be very high. Considerable confidence in the evidence for making such a change will reasonably be demanded. In another, little may be lost in making a procedural change, whereas to let matters drift may amount to bad management. The decision to change in this case does not require full confidence in the significance of the evidence for it.

Often the implications of the survey data are not perceived in such a major decisionmaking role. In simpler terms, the user may look at a table of figures, wanting an answer to the question: "Am I on the right track; are yields rising?" An answer—"Yes, probably"—which implies a better than even chance that it is so, may serve a more useful purpose than a conservative response of "Not known" made on the basis of rigorous confidence standards. Such an answer must of course be based not only on the actual increase but also consider past variability.

Sampling errors can be calculated; nonsampling errors often cannot. Yet as discussed earlier, these may be crucial in any consideration of the accuracy of the estimates presented. The author of the tables should have detailed knowledge of the source, methods used in collection, problems in analysis, and so on. It is therefore necessary to provide an assessment of the likely biases and observational errors that may affect the results, even though such an assessment lacks the mathematical support that the calculation of sampling error commands. To state that the estimate of output is X with a standard error of, say, 4 percent, leaving unstated the suspicion that the measurements were seriously biased, is to give a totally spurious impression of accuracy. A description of the survey procedures, design, methodology, nonresponse rate, practical problems, results of field checks, and so on may serve the purpose. But because most readers may skip such introductory sections, the reporter needs to comment on the implications of these in the text which discusses the findings.

Writing the Main Report

As we have repeatedly stressed in this and the companion volume, the functions of monitoring and evaluation are not achieved by the compilation of bulky semiannual or annual reports to higher authorities—reports which go largely unread or at least unused in terms of affecting implementation procedures. Such reports may be required, and must therefore be laboriously compiled, as a record for auditing. In the proper exercise of monitoring and evaluation, however, there will be occasions when surveys and studies are conducted within the work program. The reports that give the results of such surveys should be structured to convey the following information:

- a. The topics under discussion
- b. The source and reference period of the material presented
- c. The methodology of the survey
- d. A summary of the data
- e. The necessary comments regarding the accuracy of the data
- f. A review of the highlights and implications of the data
- g. The drawing together of conclusions, options for the consideration of managers, and, as appropriate, recommendations.

The order in which these sections are arranged needs careful consideration and will vary according to the type of user addressed. Sections f and g may be given as a summary at the very beginning of the report—taking maximum advantage of the limited time the reader has to concentrate on the report. If the findings are important enough, the further attention of the reader to the body of the report may be secured. The background description contained in b and c may be relegated to appendixes or retained as a necessary introduction at the beginning of the main body of the report. Within the report itself, d may involve only a succinct presentation of tables; the more detailed tables may appear as appendixes.

Even when the author follows a structure consistent with this outline, a common fault is excessive detail. Although it is important to provide the necessary methodological background, there is rarely need to drag the reader through a blow-by-blow account of the formulas for making sample estimates or calculating averages. Excessive detail often is provided to impress the reader, perhaps because the author doubts the impression that will be made by the analysis itself and the conclusions. Cover each vital section, but as succinctly as possible.

The text which accompanies the numerical data and tables in a report is intended to supplement the tables and thus assist in the process of converting the data into information. The method of presentation of this essential supplement and the style in which it is written will vary according

to the writer; there is no harm in—and much in favor of—such individuality, as long as the ultimate objective of clearly conveying the information to the reader is achieved.

Each table needs to be accompanied by a paragraph or two of text which draws attention to the significant facts, states their likely meaning, sets them in their correct perspective in the context of the information contained in other tables, and assesses the reliability of conclusions that seem indicated. The alternative is to consolidate the text in one section and the tables in another. We do not recommend this for two reasons. First, the general reader may not refer to the tables when reading the text, so that the impressions gained are divorced from the underpinning provided by the actual figures. Second, many users of the report may turn directly to the tables to obtain figures. The text, with its insights and caveats, will not be referred to, which leads to the danger that the data will be misinterpreted. If the text and tables are integrated, these dangers are at least minimized.

The text should be short and spare in style. Arguments should be sharply presented. Side issues should be treated separately and not allowed to confuse the development of the main thesis. Brevity, however, does not release the writer from the need to give his assessment of the evidence; rather, the briefer and more aggregated the presentation, the more important his role in guiding the reader to the correct interpretation. Highly aggregated summary tables cannot be investigated by the reader; it is therefore important that the writer provide the necessary explanatory comments.

The reader of any report is asked to trust that the reporter has been objective and frank in marshaling the facts. It may be assumed that deliberate falsification has not been introduced, but selective omission of contradictory evidence, removal of data from their appropriate context in order to make spurious generalizations, and failure to warn regarding possible biases may produce a false picture.

It is unfortunate that much of the professional training of economists, statisticians, and social scientists does not address how to interpret a table and write a clear summary of this interpretation. Many monitoring officers fail to become more integrated into the management team in part because they lack conviction in listing the issues revealed by a report. If they are uncertain of their writing ability, they would be better off restricting the text, as far as is possible, to a simple listing of the points to be made. We ourselves have made liberal use of lists in both this volume and its companion, and we started this section on the writing of reports with an example.

Suggested Readings

- Bailey, Kenneth D. *Methods of Social Research*, 3d ed. New York: Free Press, 1987.
- Bogdan, Robert C., and Stephen K. Biklen. *Qualitative Research in Education*. Boston: Allyn & Bacon, 1982.
- Casley, Dennis J., and Denis A. Lury. *Data Collection in Developing Countries*. 2d ed. London: Oxford University Press, 1987.
- . *Monitoring and Evaluation of Agriculture and Rural Development Projects*. Baltimore: Johns Hopkins University Press, 1982.
- Cook, Thomas D., and Charles S. Reichardt. *Qualitative and Quantitative Methods in Evaluation Research*. Beverly Hills, Calif.: Sage, 1979.
- Douglas, Jack D. *Investigative Social Research*. Beverly Hills, Calif.: Sage, 1976.
- Ehrenberg, A. S. C. *Data Reduction*. London: Wiley, 1978.
- . *A Primer in Data Reduction*. Chichester, Eng.: Wiley, 1982.
- Food and Agriculture Organization of the United Nations. *Estimation of Crop Areas and Yields in Agricultural Statistics*. FAO Economic and Social Development Paper 22. Rome, 1982.
- . *Programme for the 1980 World Census of Agriculture*. Rome, 1976.
- Hoaglin, David C., Frederick Mosteller, and John W. Tukey. *Exploring Data Tables, Trends, and Shapes*. New York: Wiley, 1985.
- . *Understanding Robust and Exploratory Data Analysis*. New York: Wiley, 1983.
- Ingle, Marcus D., Noel Berge, and Marcia Teisan. *Acquiring and Using Microcomputers in Agricultural Development: A Manager's Guide*. Washington, D.C.: U.S. Department of Agriculture, 1983.
- Kish, Leslie. *Survey Sampling*. New York: Wiley, 1965.
- Kumar, Krishna. *Conducting Group Interviews in Developing Countries*. Washington, D.C.: U.S. Agency for International Development, 1987.
- . *A Manager's Guide to Rapid, Low-Cost Data Collection Methods*. Washington, D.C.: U.S. Agency for International Development, 1987.
- Maanen, John Van, ed. *Qualitative Methodology*. Beverly Hills, Calif.: Sage, 1983.
- Miles, Matthew B., and A. Michael Huberman. *Qualitative Data Analysis: A Sourcebook of New Methods*. Beverly Hills, Calif.: Sage Publications, 1984.

- McCall, George, and J. Simmons, eds. *Issues in Participant Observation*. Reading, Mass.: Addison-Wesley, 1969.
- Mosteller, Frederick, and John W. Tukey. *Data Analysis and Regression*. Reading, Mass.: Addison-Wesley, 1977.
- Murphy, Josette, and Leendert Sprey. *Introduction to Farm Surveys*. ILRI Publication 33. Wageningen, Netherlands: International Institute for Land Reclamation and Improvement, 1983.
- . *Monitoring and Evaluation of Agricultural Change*. ILRI Publication 32. Wageningen, Netherlands: International Institute for Land Reclamation and Improvement, 1982.
- O'Muircheartaigh, C. A., and C. Payne, eds. *The Analysis of Survey Data*. London: Wiley, 1977.
- Patton, Michael Quinn. *Qualitative Evaluation Methods*. Beverly Hills, Calif.: Sage, 1980.
- . *Utilization-Focused Evaluation*. Beverly Hills, Calif.: Sage, 1978.
- Poate, C. D., and Dennis J. Casley. *Estimating Crop Production in Development Projects: Methods and Their Limitations*. Washington, D.C.: World Bank, 1985.
- Scott, Chris. *Sampling for Monitoring and Evaluation*. Washington, D.C.: World Bank, 1985.
- Slade, Roger H., and J. Gabriel Campbell. *An Operational Guide to the Monitoring and Evaluation of Social Forestry in India*. Rome: Food and Agriculture Organization of the United Nations, 1987.
- Smith, Robert N., and Peter K. Manning, eds. *Qualitative Methods*. Volume 2 of *Handbook of Social Science Methods*. Cambridge, Mass.: Ballinger, 1982.
- Spradley, James P. *The Ethnographic Interview*. New York: Holt, Rinehart & Winston, 1979.
- . *Participant-Observation*. New York: Holt, Rinehart and Winston, 1980.
- Stewart, David W. *Secondary Research: Information Sources and Methods*. Beverly Hills, Calif.: Sage, 1984.
- United Nations Department of International Economic and Social Affairs, Statistical Office. *Handbook of Household Surveys*. New York: United Nations, 1984.
- Yin, Robert K. *Case Study Research: Design and Methods*. Beverly Hills, Calif.: Sage, 1984.

Index

- Adoption rates for project services and inputs, 1, 3, 57, 59, 131, 134
- Air or ground transects, 99–100
- Analysis of data. *See* Data analysis
- Area measurement. *See* Crop area measurement
- Baseline surveys. *See* Socioeconomic baseline surveys
- Beneficiary of project. *See* Project beneficiary/participant
- Bias, 51–52, 77, 81–82
- Causes and effects of events, 2–3
- Class intervals. *See* Frequency distributions
- Cluster sampling, 80, 92–93
- Community interviews (CIs): balancing participation in, 30–31; and data collection, 31–33; defined, 26; interview teams and, 30; limitations of, 40; postmeeting conversations and, 33–34; size and timing of, 29; structured guides for, 28–29
- Computerization of data collection, 8–9
- Confidence levels, 84, 86, 92
- Crop area measurement, 96–98; by air or ground transects, 99–100; by grid sampling, 100; within holdings, 100–102; by point sampling, 100
- Crop cutting, 102–09
- Crop production and yield measurement, 96–99; biological yield and, 98; crop cutting method of, 102–09; economic yield and, 98; eye estimate assessments and, 102, 113; farmer estimates and, 1, 102, 111–13; harvested yield and, 98; harvesting method of, 102–03; harvest unit sampling and, 108, 110
- Cross-sectional surveys, 58
- Cross-tabulations: and data presentation, 151, 155–58; in statistical analysis, 137–38
- Data analysis, 7–9; graphic, 115–17. *See also* Exploratory analysis; Statistical analysis
- Data collection: cause and effect and, 2–3; community interviews and, 31–33; computerization of, 8–9; constraints on, 6–7; data requirements for, 57; descriptive, 2–3; explanatory, 2–3; predictive, 2–3; purposes of, 2–3. *See also* Qualitative data collection; Quantitative data collection
- Data presentation: accuracy and significance of data for, 161, 163, 165–66; cross-tabulations and, 155–58; frequency distributions and, 149–50; graphic, 149, 161–63, 164, 166; percentages and, 159–60; principles of, 148; tabular, 149; techniques for, 149; trunca-

- tion of digits and, 150, 154–55.
See also Report of survey
- Diagnostic studies for monitoring, 82
- Document summary sheets, 49–51
- Enumerator: role of, 74–75; training of, 7, 59
- Errors. *See* Nonsampling errors; Sampling errors
- Estimator, defined, 76
- Exit polls, 90–91
- Exploratory analysis: adoption rates and, 131; dispersion measurement and, 126–27; fitting straight lines and, 127, 128; graphic examination and, 115–17; moving averages and, 129–30; ordering data for, 117–23; purpose of, 114–15; residuals and first differences and, 127–29, 130; subsample analysis and, 131; transformation of data and, 123–26
- Farmer. *See* Holder
- Farmer estimates, 1, 102, 111–13
- Focused group interviews (FGIs): controlling, 38–39; defined, 26–27; duration of, 36; guides for, 34; limitations of, 40; location of and seating for, 36; opening statements in, 36–37; participant selection for, 35–36; questions for, 37–38; recording of, 39–40; size and composition of, 34–35
- Formal sampling. *See* Probability sampling
- Frequency distributions, 149–50
- Graphic analysis of data, 115–17
- Grid sampling, 100
- Ground or air transects, 99–100
- Group interviews. *See* Community interviews; Focused group interviews
- Harvest unit sampling, 108, 110
- Holder, defined, 60–61
- Holdings: area measurement of, 100–01; defined, 61
- Household, defined, 60
- Household surveys. *See* Socioeconomic baseline surveys
- Informal sampling: defined, 78; disadvantages of, 81; probability sampling and, 78–83
- Informants. *See* Key informants
- Interviews, 74–75. *See also* Community interviews; Focused group interviews; Qualitative interviews
- Key informants, 24
- Longitudinal surveys, 57–58
- Mixed cropping: defined, 62; production measurement of, 98–99
- Nonsampling errors, 84–85, 145–47
- Observation record forms, 49–50
- Panel surveys, 57–58
- Parcel, defined, 61
- Participant observation: advantages of, 41–42; conceptual framework for, 42–45; defined, 41, 41n; fieldwork guidelines for, 46–49; limitations of, 52–53; minimizing biases in, 51–52; as qualitative data collection method, 5; site selection and timing for, 45–46; study instruments for, 49–51
- Plots: and area measurement, 100–01; defined, 62
- Point sampling, 100
- Probability sampling: characteristics of, 78; and informal sampling, 78–83; objections to, 79–80
- Project beneficiary/participant, defined, 62–63
- Purposive sampling, 78, 78n, 79
- Qualitative data collection, 3–6
- Qualitative interviews: controlling conversations in, 19; guidelines

- for, 15–21; informal, conversational, 11–12; initial contact for, 15; key informants and, 24; limitations of, 24–25; neutral attitude in, 19–20; potential of, 10; probing in, 18–19; question sequencing in, 15–16; recording of, 20–21; reliability of, 21–24; role playing in, 17–18; semistructured, open-ended, 13–14; topic-focused, 12–13
- Quantitative data collection, 3–6, 31–33. *See also* Structured surveys
- Questionnaires: leading questions in, 68–69; open-ended versus closed questions in, 64, 67; pretesting of, 73–74; question sequencing in, 69–70; recording of, 71–73; response set and, 69; and socioeconomic baseline surveys, 55, 56; tabular format of, 65; verbatim or shorthand questions in, 70–71; wording of, 67–68
- Quota sampling, 78
- Random sampling, 77, 78n
- Random selection of respondents, 5–6, 78n
- Recall period, defined, 63–64
- Reference period, defined, 63–64
- Report of survey: document summary sheets and, 49–51; observation record form and, 49–50; study instruments and, 49–51; written, 20–21, 49–51, 167–68
- Sample, defined, 76
- Sample design: constraints on options for, 6–7; costs of, 91–92, 93; random selection for, 5–6; selection process for, 77; single-stage, 85–92; two-stage, 92–94
- Sample size: for comparison of two rates, 91; and sampling errors, 83–85; for two-stage sampling, 93–94
- Sampling: bias, 77, 81–82; cluster, 80, 90–92; for comparison of two rates, 91–92; disadvantage of, 80; fractions and, 78, 91–92; frame, 77, 80; grid, 100; harvest unit, 108, 111; informal, 78, 81; listing and, 80–81; point, 78; probability, 78–83; purposes of, 76; purposive, 78, 78n, 79; quota, 78; random, 5–6, 77, 78n; for rare events, 94–95; scientific, 78, 89; stratified, 79, 87–89, 91; systematic selection and, 78, 78n, 87; two-stage techniques for, 92–94; variance, 77, 83, 83n, 86, 89
- Sampling errors: calculating, 165–66; defined, 77; sample size and, 83–85
- Socioeconomic baseline surveys, 3, 54–56, 56n
- Statistical analysis: for comparing two sample means, 133–34; crosstabulations in, 137–38; dangers of, 132; regression analysis and, 127, 128–29, 142–47; 2 x 2 tables and, 134–37; variance analysis and, 138–42
- Stratified sampling, 79, 87–89, 91
- Structured surveys, 1; advantages of, 4; concepts and definitions for, 54, 59–64; cross-sectional, 58; data requirements for, 57; design of, 4; interviewing respondents for, 74–75; longitudinal, 57–58; panel, 57–58; planning for, 56–57; questionnaire construction for, 64–74; role of enumerator in, 74; single-round, 58–59. *See also* Socioeconomic baseline surveys
- Survey design: constraints on, 6–7; longitudinal, 57–58; socioeconomic baseline, 56n; structured, 4
- Survey, report of. *See* Report of survey
- Systematic selection, 78, 78n, 87
- Tables, 149; as format of questionnaire, 65; 2 x 2, 134–37
- Transects, 99–100
- Universe: defined, 76; requirement

- for knowledge of, 80
- Written report of survey. *See* Report of survey
- Variance: analysis of, 138–42; sampling, 77, 83, 83n, 86, 89
- Yield measurement. *See* Crop production and yield measurement

THE MOST RECENT WORLD BANK PUBLICATIONS are described in the catalog *New Publications*, which is issued in the spring and fall of each year. The complete backlist of publications is shown in the annual *Index of Publications*, which is of value principally to libraries and institutional purchasers. The latest edition of each is available free of charge from Publications Sales Unit, The World Bank, 1818 H Street, N.W., Washington, D.C. 20433, U.S.A., or from Publications, The World Bank, 66, avenue d'Iéna, 75116 Paris, France.

A Joint Study



The World Bank



International
Fund for Agricultural
Development



Food and Agriculture
Organization
of the United Nations

THIS VOLUME PROVIDES SIMPLE PRACTICAL METHODS of collecting and analyzing data for monitoring and evaluating agricultural projects. It further explains the techniques referred to in the companion volume, *Project Monitoring and Evaluation in Agriculture*. The authors have selected methods that are both simple and inexpensive because of the limited resources of many development projects and because such methods are usually preferable.

Each chapter deals with a specific area of data collection, analysis, and use. The subjects covered are qualitative and quantitative methods of data collection, structured surveys, sampling problems, crop measurement, preliminary and exploratory analysis of data, formal analysis, and data presentation.

Emphasizing qualitative interviewing methods, the book explains how to design qualitative surveys and how to use methods of participant observation. It also extensively describes various aspects of quantitative methods for gathering and analyzing data, such as sample theory and selection, but treats as well the possible pitfalls of quantitative methods and when these techniques may not be appropriate. Together with its companion volume—which explains in detail the concepts of monitoring and evaluation—this book will be useful for those who design and implement these systems and as texts for regional and national training programs. This book and its companion are a joint effort of the World Bank, the Food and Agriculture Organization of the United Nations, and the International Fund for Agricultural Development.

DENNIS J. CASLEY is chief of the Operations Monitoring Unit, Central Operations Department, the World Bank. He wrote this book while a staff member of the Bank's Agricultural and Rural Development Department. KRISHNA KUMAR, a senior analyst at the Center for Development Information and Evaluation, U.S. Agency for International Development, was a consultant to the World Bank for the writing of this book.

Cover design by Bill Fraser

The Johns Hopkins University Press
Baltimore and London

0-8018-3669-7