Monitoring and evaluation of agricultural change

Josette Murphy and Leendert H. Sprey

International Institute for Land Reclamation and Improvement/ILRI
P.O. Box 45, 6700 AA Wageningen, The Netherlands 1982.

The work of Dr. Murphy has been financed by the Upper Volta Office of the United States Agency for International Development under contract PSC 686-092-080.

ISBN 90 70260 743

Printed in The Netherlands.

PREFACE

Since 1976, we have been involved in the monitoring and evaluation of agricultural change, mostly in the semi-arid tropics, Josette Murphy as an economic anthropologist and Leendert Sprey as an agricultural economist. While training African staff, we felt that a book of guidelines on how to design and conduct a monitoring and evaluation program was needed. The book would be adapted to the agricultural practices of subsistence farmers in the semi-arid tropics, and easy to use by evaluators with little experience in survey work and limited resources to do the job.

We are very grateful to the International Institute for Land Reclamation and Improvement (ILRI), Wageningen, the Directorate General of International Cooperation (DGIS), The Hague, and to the United States Agency for International Development (USAID), Washington, for sponsoring the preparation and publication of this book, both in English and in French.

Many reviewers have commented on our drafts. We especially wish to thank experts in the following institutions:

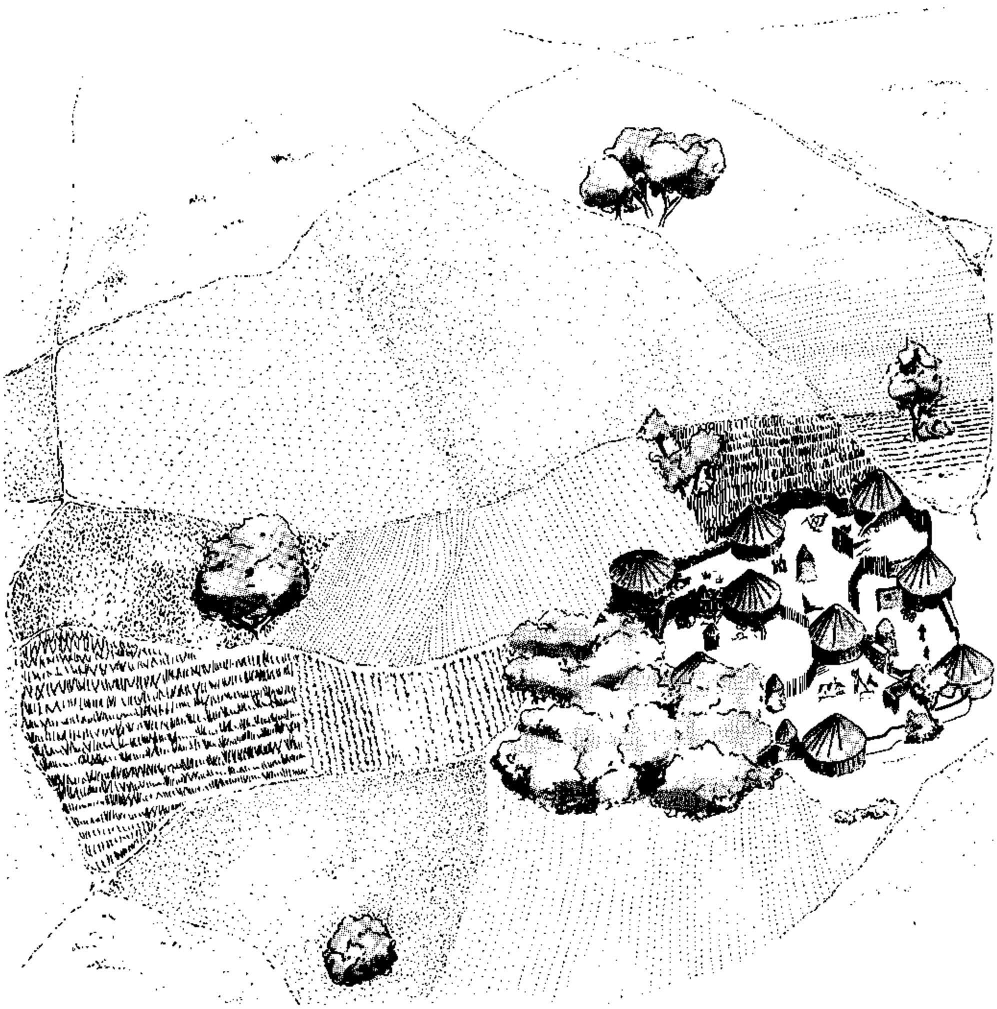
- Directorate Agricultural Research (DLO);
- Hunting Technical Services Ltd.;
- Institut National de la Statistique et des Études et Économiques (INSEE);
- International Centre for Corn and Wheat Improvement (CIMMYT);
- International Crops Research Institute for the Semi-arid Tropics (ICRISAT);
- International Institute for Aerial Survey and Earth Sciences (ITC);

- International Institute for Land Reclamation and Improvement (ILRI);
- Royal Institute for the Tropics (KIT);
- United States Agency for International Development (USAID);
- United States Department of Agriculture (USDA);
- University of Groningen (RUG);
- Wageningen Agricultural University (LH);
- World Bank.

Peter Matlon deserves a special acknowledgment for his detailed comments and for allowing us to test our draft with his enumerators.

We also wish to thank Margaret Wiersma-Roche, our editor, and Ria Beerens, our typist, for their thorough and accurate work over many months. Our appreciation also goes to Joop van Dijk for his drawings and to Johan van Manen for his graphs.

CONTENTS



1 INTRODUCTION

1.1 The need for information

Small farmers in the semi-arid tropics have been making an efficient use of their limited resources for centuries, but as populations have increased, their traditional practices have become inadequate. In many countries where agriculture is the major source of income, efforts are now under way to increase agricultural production by improving farming practices and conditions.

The planners and policy-makers in these countries are in a better position to succeed in developing their agricultural sectors if they are well informed about existing farming conditions and about the impact of development activities as they are occuring. More specifically, they need to know about farming conditions and farm results before, during, and after project implementation, and they need to be able to compare these conditions and results with those prevailing elsewhere in the country.

Monitoring and evaluation - a system of data gathering, analysis, and feedback of farm information during the life of a project - can assist development efforts in two ways:

- It can allow project managers to adjust their activities to the farmers' needs and constraints;
- It can provide planners and policy-makers with up-to-date information on agricultural change.

To be done successfully, monitoring and evaluation must meet three conditions:

- The program of data gathering and analyses must fit the specific needs of the institutions or specialists that will use the information;
- The methodologies applied must be appropriate to the personnel and material resources available;
- The results must be presented in a format and at a time convenient for planning and management decisions.

1.2 Purpose of this book

Few people in developing countries are trained or experienced in organizing the monitoring and evaluation of a project. Even if they are, working with subsistence farmers in the semi-arid tropics is likely to be difficult. Staff are likely to need training, and will probably have to be closely supervised. Problems are bound to arise with transport and materials. Conditions are likely to be such that only fairly imprecise data can be obtained, which prevents the use of advanced statistical techniques. Trying to work to rules set for a different environment will probably lead more to frustration than to results.

In this book, we have elaborated a program for monitoring and evaluation that fits the conditions in the semi-arid tropics. It is not a review of existing methodologies, but is a tool that can help the reader to select a course of action appropriate to his situation.

The book has been written for those who will be called upon to design a program of continuous data gathering and analysis and to supervise the implementation of that program, whether for a project or for a government agency. We assume that our readers have a good general education, but need not have any previous knowledge of agronomy, economics, or statistics.

1.3 How to use this book

This book is focussed primarily on subsistence agriculture in the semi-

arid tropics, where family farms grow the crops they need for auto-consumption under dry-land conditions. It assumes that farming practices include shifting cultivation and the use of manual labour, most of it supplied by the family itself. Development projects in such areas are likely to introduce the use of animal traction, fertilizer, and improved varieties of crops.
Most of the examples cited are from the semi-arid areas of West Africa. The book is applicable elsewhere, as long as agricultural conditions are comparable. Much of the general information on monitoring and evaluation may also be useful in other types of projects (e.g. those that include irrigation) or those in other climatological zones.

The book is presented in two parts. Part 1, *General Principles*, describes, step by step, how to organize a monitoring and evaluation unit. Part 2, *Methodologies*, is a detailed elaboration of the activities described in Part 1; it also discusses the problems likely to occur, so that the person called upon to organize a monitoring and evaluation unit will be forewarned.

A companion volume, *Introduction to farm surveys*, has been written to help the director of a monitoring and evaluation unit train his enumerators. It is divided into lessons, beginning at a very simple level and becoming progressively more difficult. At the end of each chapter is a set of exercises or questions, the solutions to which are given at the back of the book. The subjects covered include surveys, basic statistics, measurement techniques, agricultural economics, and some basic agronomy. The enumerators' supervisors, who are assumed to have a slightly higher level of education and some experience in data gathering, may also benefit from reading that book.

2 THE CONCEPT OF MONITORING AND EVALUATION

2.1 Introduction

Our understanding of farming in the semi-arid tropics is not yet thorough enough for development specialists to recommend improvements that will guarantee more food and a better way of life for the small farmers. An agricultural development project is still very much a trial-and-error process, so its managers and specialists need a systematic feedback of information from the farmers for whom the project is being implemented. To provide project personnel with the information they need to understand the results of their activities - at a time when they can still modify the project program - data should be gathered continuously and analyzed without delay.

A program combining such data gathering, analysis, and interpretation is called a monitoring and evaluation program.

- Monitoring means recording the activities of the project staff and the farmers and measuring the results that are achieved;
- Evaluating means interpreting the results to determine whether these activities have had the results that were expected and, if not, why not.

The combination of monitoring and evaluation goes beyond the simple keeping track of events; it makes it possible to understand why events occur.

2.2 Different roles of monitoring and evaluation

2.2.1 Role in project implementation

Monitoring and evaluation provide a concise, permanent flow of information on agricultural practices actually being applied and the results actually being achieved. Monitoring is often organized within the project itself, so can be adapted to the needs of the project's specialists and managers. When the farmers' activities are monitored in this way, project specialists can evaluate whether the farmers' situation is improving as expected and, if not, they can try to find out why. They can then recommend corrective action by strengthening or modifying one or more components of the project. Project managers who have the results of combined monitoring and evaluation at their disposal can adjust their programs early enough to minimize detrimental effects before these become major problems. In this way, the project is able to bring a program well adapted to farming conditions - one that has a better chance of succeeding.

Monitoring and evaluation in an agricultural development project are thus management tools that serve as an 'early warning system', making it possible to adjust implementation procedures and modify technical recommendations whenever needed. They may even lead to changes in project objectives and priorities - should this become desirable - at any time during project implementation.

2.2.2 Role in planning future projects

Monitoring and evaluation provide the type of information that makes a comprehensive ex-post evaluation of a project possible. They also allow comparisons between project farms and traditional farms or between the farms of one project and those of another. Each project becomes a learning process through which valuable experience is accumulated for future projects, thereby providing a better understanding of the conditions under which development efforts are likely to succeed and giving planners a sound basis on which to work.

2.2.3 Role in agricultural research

Like project staff, specialists working in agricultural research need information on farmers' activities. It is increasingly accepted that, before improved farming techniques can be recommended for inclusion in extension programs, they should be adapted to local farming systems and tested under real farming conditions. At national and international agricultural research centres, efforts are now under way to develop technical packages that are better adapted to the needs and capacities of small farmers. This approach requires detailed information on traditional farming systems and on the processes through which the natural, social, and economic constraints are likely to influence future change. Essential for such problem-oriented research is a systematic continuous feedback of farmers' responses to extension programs and of on-farm testing of proposed technology.

2.3 Administrative structures for monitoring and evaluation

The monitoring and evaluation of farm results can take place within different administrative structures. The three most likely structures are:

- A specialized unit within a large project;
- An informal grouping of extension agents in a small project;
- A national or regional institution.

2.3.1 Monitoring and evaluation in a large project

The systematic feedback required by managers and specialists makes it worthwhile for a large project to create a monitoring and evaluation unit of its own. This unit conducts programs of data gathering and analysis as requested by project personnel. Ideally, the unit should be directly under the authority of the general management of the project, but it is often found within its planning or economic division.

2.3.2 Monitoring and evaluation in a small project

In a small project, where a separate unit would be too expensive, monitoring may be done by extension agents. The data they gather are then analyzed by the project's agro-economist. Done this way, the monitoring will usually be fairly superficial. If extension agents have the dual task of extension and data collection, they will only be able to spend a small part of their time on data collection and the resulting evaluations will be limited to the practices being promoted by the extension programs.

2.3.3 Monitoring and evaluation in national or regional institutions

Good data on local farming systems and their potential for improvement are also needed by policy-makers at regional and national levels. An institution like the National Bureau of Statistics or the Ministry of Agriculture could set up coordinated monitoring and evaluation in local agencies such as regional extension services, agronomic research centres, and individual projects.

If such an institution undertakes the monitoring and evaluation of projects in a region or a country, it can include subjects that might not be of direct interest to managers of individual projects but which are of regional or national importance. The institution will be able to evaluate the impact of a project on the region or the country, something project managers are seldom able to do. Reports from such an institution will be more appropriate for regional and national planning than uncoordinated reports from individual projects could ever be.

In the same way, the efforts of regional development agencies can be more effective if there is a coordinated body of data available so that the results of each agency can be evaluated and compared with those of other development agencies in the country.

Other advantages of institutionalized monitoring and evaluation are that baseline studies can be conducted on the agricultural practices in a region. Such studies can provide descriptions of local farming practices and achievements, can identify the difficulties encountered by farmers, and can point out the need for further research or assistance. For effective national and regional planning and for project identification, such studies are essential.

Monitoring and evaluation within the three structures described above may differ in the type of data collected, but all will follow the same procedures of selecting topics to be monitored (the variables) and conducting surveys. For ease of presentation, this book will assume that the monitoring activities are conducted by a separate unit in a large development project, under the supervision of a unit director, and that the results will be evaluated by the unit director and project specialists together. The procedures described can be followed in smaller projects as well, although it is likely that fewer variables will be monitored and that surveys will be conducted on a smaller scale and perhaps with less intensive survey schedules. A national or regional monitoring and evaluation institution can also follow the procedures described, although its surveys will be done on a larger scale and the results will be aggregated at the regional or national level.

3 DELIMITING THE WORK OF A MONITORING AND EVALUATION UNIT

The key to usefulness - the standard of success achieved by a monitoring and evaluation unit - lies right at the beginning when the work of the unit is being delimited. What matters is that the unit provides information which is relevant for its users and that the information is valid and is presented at the right time. Difficulties and delays inherent to data gathering and processing can be kept to a minimum if, right at the beginning, *in consultation with the users*, it is decided what data are really pertinent and necessary.

3.1 Identifying the users

A monitoring unit exists to serve the people who request information. These users of information will have diverse backgrounds and will fulfil different roles in a project. Potential users may be:

- Project managers, who need continuous information to be able to plan ahead, and who may wish to redefine the strategies of the project at regular intervals;
- Agronomists, who wish to orient their research in line with the limitations and constraints of small farmers;
- Extension service, which wants to adapt its extension program to the achievements of extension agents and farmers;
- Agricultural credit section, which may want to adapt its repayment policy to farm results;
- Economists, who want to measure the economic impact of the project on

the farmers;

- Sociologists, who want to know of any socio-economic constaints to change and of the social impact the project is having, so that they can coordinate their studies with the agro-economic surveys.

The information obtained through monitoring and evaluation within a project can also be of interest to government planners and to agencies outside the project, but as these are not the primary users of the information, they will usually not participate in delimiting the scope of work of the unit, as users within the project will.

Consultation with the users

3.2 Defining what information is needed

3.2.1 Consultation with the users

Users often have difficulty in formulating clearly what type of information they want from the unit. The unit director should therefore take the initiative to contact potential users and find out what kind of information would be most useful to them. During this consultation, the

unit director can help the user define his needs by explaining what data pertaining to the user's discipline can be gathered by the unit and how the data can be analyzed and reported.
At that stage, a user is likely to express very general wishes such as 'Find out if the project farmers are achieving the expected farm results'. But this is too vague. For each topic, the user should be asked to specify the following:

- What information is needed?
- How will the information be used?
- What level of accuracy is required of the data?
- What population is to be studied?
- What analysis is required?
- What format should the report have and when is it required?

Making each user explain the information he requires and how he intends to use it will help him become more specific in his wishes. The user quoted above, who wants to know about 'farm results', may have in mind an evaluation comparing the average yields of each crop in each village with the yields that are expected as a result of the technical improvements being promoted. Or, he may want an evaluation comparing the farmers' income per capita with the national average income. These are two different objectives which call for different approaches in data gathering and analysis.
Most of the evaluations performed by the unit will involve comparing farm data with some kind of reference, such as the expected results listed in the project papers or the results obtained by farmers outside the project. It is important that the users identify the references that are to be used, or that they approve references proposed by the unit. During the discussion on the information required, the users, who are better aware of the specific factors that need to be considered for a correct analysis, should make these factors clear.

The level of accuracy required of the data is mainly determined by the type of evaluation to be done. The accuracy that can be attained depends upon the capacity and resources of the unit. Users have a tendency to ask for data of greater precision than they really need. It is therefore in the interest of the unit to discuss this matter with the users so

that the unit's time and manpower will be used as efficiently as possible.
The population the unit will study is defined as the aggregate of elements which are the object of study. Those elements can be individual persons, but also things like farms, fields, or animals.

For example: If the farming practices in the project are to be studied, the population consists of all the farms in the project. If the production of cowpeas is to be studied, the population consists of all the fields of cowpeas in the project.

Users will frequently ask the monitoring unit to draw conclusions about an entire population in the project area. This population may then have to be divided (i.e. stratified) according to agro-climatological zones and perhaps to other differences important to the user. Sometimes a user is not interested in the entire population of the project but only in a certain segment of it.

For example: He may want to learn about potential income from petty trade and needs to find out how trading activities can be developed. If so, the population of active traders will be studied.

The users' requirements as to population must be known by the monitoring unit when putting its program together. If many different populations are to be studied, the unit may find itself overloaded with work. It can fulfil only a limited number of requests at a time. Usually therefore not all the elements of the population are studied but only a sample chosen to represent the population. The criteria for the selection of a sample should be discussed with the user of the information.
The type of analysis requested by the users is likely to be fairly simple and will include averages, frequencies, measures of dispersion, and simple comparisons between groups of the population. If an interesting trend is noticed during analysis, it should be pointed out to the user, although, in general, the user should be aware that he will receive only what he asks for, as it is no use performing additional analyses that nobody will use. Further analysis can be done later on request, if the data base make it possible.

The way in which results are presented in a report depends not so much on the type of information, but on the user's preferences. If he desires a numerical or graphic presentation of results, these wishes should be respected as far as possible. Instructions for lay-out of the text and tables, and to whom the reports should be sent and when, should also be obtained in advance.
A reporting schedule, including periodicity and deadlines, should be arranged before the unit's program of work is prepared. Rapid reporting is possible if simple analyses only are required. A reporting schedule in two stages may be useful, with frequent preliminary reports and a more in-depth evaluation report after each crop season.
It may not always be possible to obtain clear, specific requests from a user, but only a rough list of requirements and how the results should be analyzed. The unit director can then make a more detailed list, covering the six points outlined above, and submit this list for the user's approval.

3.2.2 Listing the data required

Preliminary discussions with potential users enable the unit director to make a provisional list of all the information that the unit will need to collect. He should find out whether any of this information is already available or whether it can be expected to become available from other sources. He should also remember that he might need information not explicitly requested by the users but necessary for the analysis.

For example: For an analysis of yields, he may require a detailed description of the farming techniques being promoted by the project, although this has not been requested.

A list of the data to be collected by the monitoring unit and of the analyses it will be required to make can now be prepared. Examples of possible lists for users from various disciplines are given in the following pages. These examples are meant merely to show how to go about making such lists; they are not standard programs to be used in real projects.

Example A: Request from extension service

Data required: Main points of the extension program and actual realization in the field (program actually promoted is seldom identical with official program), the reason why the farmers do or do not follow the program (if possible, data on farm income).

Use of data: To determine the extent to which the farmers adopt the extension program, for a better adaptation of the extension program to farmers' needs.

Level of accuracy: This survey seeks only to identify problem areas so the data need not be detailed.

Population to be studied: All farmers in the project area, with special attention to the areas where the program does not seem succesful. A sample representative of the farmers in the project area can be selected.

Analysis to be done: Calculation of means, frequencies of observations, classification of results according to groups of farmers with the same kind of attitude towards the extension program (comparison of farm income in different groups if income data are available).

Reporting format: As far as possible, graphic presentation per extension region.

Reporting time: Once a month throughout the agricultural season and overall report later.

Example B: Request from sociologist

Data required: Inventory of group activities during the last crop cycle.

Use of data: To identify the types of group activities undertaken by the farmers in order to investigate the possibilities of using existing group activities to introduce new soil management techniques.

Level of accuracy:	Inventory by categories (determined by sociologist) of type of activity, group size, and composition.
Population to be studied:	The entire population of adult people in the project (not only farmers) from which a representative sample can be selected.
Analysis to be done:	Calculation of the frequency of each type of activity as percentage of population.
Reporting format:	Tables and graphs with frequencies, and a discussion of the results.
Reporting time:	No time limit.

Example C: Request from manager

Data required:	Yields and areas cultivated, or production figures and areas cultivated, or yields and productions, for all crops.
Use of data:	Data on the farmers' agricultural production will be presented to the Government and funding agencies to illustrate the project's achievements.
Level of accuracy:	Rough figures are sufficient, with yields rounded off to 100 kg/ha.
Population to be studied:	All project farmers; for each crop a representative sample can be selected, with the possibility of selecting a different sample for each crop.
Analysis required:	For each crop, average yield and its distribution in different locations, and overall average yield and its distribution, yield, and total production.
Reporting format:	Tables of results and graphic presentation of their distribution.
Reporting time:	December, in the annual report on the crop season.

Example D: Request from manager

Data required:	For each farmer studied, the following data: - Yield and area cultivated, or productions and area cultivated, or yields and productions of each crop cultivated; - Size of the household; - Quantities of inputs used (fertilizer, seed, insecticides etc.); In addition to farm data, the following general data: - (Market) prices of inputs and crops cultivated; - National farm income (from national or international publications).
Use of data:	The progress of the project is to be measured by farm income from agriculture per capita, which will be compared with national agricultural income level. Results will be presented to Government and funding agencies.
Level of accuracy:	Rough results are good enough (e.g. yields rounded off to 100 kg).
Population to be studied:	All the farmers of the project. A representative sample of the project population of farmers can be selected.
Analysis to be done:	Calculation of farm income per capita, average and distribution of income for different locations in income classes.
Reporting format:	Numbers, comparing farm income in the project with national income (per capita), including an explanation of the factors influencing income.
Reporting time:	December, to be included in the annual report on the crop season.

Example E: Request from agronomist

Data required:	For each sorghum field studied: the yield, fertilizer use, timing, and method of application, timing and quality of agricultural operations, damage to sorghum fields, soil type, and rainfall figures.
Use of data:	The response of sorghum production to fertilizer under farm conditions is needed for a better adaptation of trial fields to real farming conditions.
Level of accuracy:	Yield, fertilizer use as accurate as possible. Other, more qualitative data need not be very precise as they will be used more to group farmers into technological categories than to explain individual yields.
Population to be studied:	One location is to be selected for this study. This location can be a village or a small region, which can be considered more or less representative of the entire project. Here, the population consists of all the sorghum fields in this location. A random sample of sorghum fields will be drawn according to the stratifications identified by the users.
Analysis to be done:	Calculation of means, variances, and regressions.
Reporting format:	Compatible with agronomic reporting.
Reporting time:	1-2 months after harvest.

Example F: Request from economist

Data required:	Information is needed about all sources of income for each member of the household (income from agriculture, trade, livestock, etc.) and the costs incurred to achieve that income, number of people living on the farm and their nutrition requirements, market prices of products, obligatory expenses (taxes, household equipment etc.).
Use of data:	To measure household income from all sources, needed to identify sources and amount of disposable income.
Level of accuracy:	Data should be as precise as possible, but theoretical nutrition requirements can be used.
Population to be studied:	The total population of the project. Data can be gathered on one or more different samples, representative of the entire population.
Analysis to be done:	Calculate total net income from each activity and compare this with the cost of the estimated food requirements of the household.
Reporting format:	Distribution of disposable income per region presented in graphs. Also means and variances are to be presented.
Reporting time:	After harvest, as soon as possible, probably 4-6 months.

From these examples, it can be seen that some data can be used to satisfy more than one request.

For example: The yield estimates requested by the agronomist (Request E), who wants to know the response of sorghum to fertilizer, can also be used, along with additional data, for the economist who requested an estimate of agricultural income (Request F).

One should be careful, however, when planning to use results of one survey with those of another survey. It is possible that the two requests concern different populations.

For example: If a survey is done on the potential of petty trade and information is gathered only from those people who are active traders, this information cannot be used to estimate the income from trade in a request like F, which wants income data on the entire population.

3.3 Agreeing on the delimitations

After the requests from different disciplines have been translated into lists, these lists can be put together to form a survey program. Taking the resources of the unit into account, the unit director designs a program in which he attempts to satisfy all requests, although in some cases, he may have to eliminate some of the desired data or decrease their desired level of accuracy.
A meeting is then held with all those who have requested information. During this meeting, the unit director will describe the extent of data gathering and analyses that can be handled, checking to see whether the different requests have been correctly understood and whether the simplifications that have had to be made because of limited resources are acceptable to the users.
It may happen that when a user first requested information, he asked for data which, given the staff and equipment available to the unit, will be difficult to obtain at the right time or at the desired level of accuracy. Obtaining these data will either overtax the resources of the unit

or entail the sacrifice of other desirable surveys. This should be explained to the user and together he and the director should try to find a solution: by postponing the inquiry, by providing additional resources to the unit, or by simplifying the survey. In such a discussion, it often appears that the data can be collected by less demanding surveys without reducing the usefulness of the data for the user. If not, the director must propose alternative ways of collecting the data, specifying the cost (in money and time) of each alternative.

For example: As a matter of course, managers often request a detailed study of agricultural labour input, with lists of numbers of hours per operation. For a unit working with illiterate farmers, this would require frequent observations or interviews throughout the cropping season, which would severely limit the number of farmers that could be surveyed and would greatly increase the amount of paperwork and tabulations required. Even if such a survey is conducted, its results are seldom accurate and will rarely be significant for policy making. So, only if labour is a major constraint in a project will it be worth the trouble of trying to gather such data. Usually, a less detailed labour survey on man-days per operation will suffice without loss of useful information.

The program agreed upon by the director and users together should not be too ambitious, especially during the first year. Material problems and delays are unavoidable even in the best of circumstances. In the beginning, inexperienced staff will not be very efficient and time has to be allowed for training. As a rule of thumb, the work should be limited to about 60 per cent of what could be accomplished under ideal conditions.

The program should have a built-in flexibility, with a core of basic data collection, and topical surveys added when required. Some time should be reserved in the program for additional, limited surveys, which may become necessary because of new developments or because unexpected trends were noticed when the first surveys were analyzed.

4 ALLOCATING THE RESOURCES OF THE UNIT

When the work of the unit is being organized, it is vital that the unit's resources be properly allocated to cover every stage of the work: data gathering, processing, analysis, and reporting. Right from the start, each stage should be properly scheduled. This applies for all the unit's resources: staff, funds, equipment, and time. In many surveys, most of the time and money is spent on data gathering, under the assumption that data processing and analysis will follow automatically. This is a dangerous fallacy because data processing is likely to be the worst bottleneck in the unit's activities. Ample time and funding must be allotted for data processing; otherwise data will pile up and much will remain unused.

4.1 Staff and budget

A monitoring and evaluation unit is likely to employ four categories of staff:

- A unit director, who delimits the work of the unit and coordinates its implementation, from preparing the surveys to reporting the results of the analyses. He supervises the staff, and arranges their training;
- Field supervisors, who coordinate and supervise the work of the enumerators and check the validity of the data. In a small unit, the unit director may handle much of the field supervision himself;
- Enumerators, who gather the data in the villages;

- Data processors, who process and analyze the data as instructed by the unit director. In a small unit, they may also handle all office work.

Funds are needed to cover the salaries and social benefits of the staff, the equipment they need, recurring expenses for paper and office supplies, transportation costs, and data processing and analysis.

The cost of a monitoring unit depends very much upon the extent of its work and the geographical environment in which its work has to be done. Many projects for which figures are available allocate about one per cent of their total cost to monitoring and evaluation. Even a unit conducting an intensive monitoring program will not cost much more than this, but will provide information that enables the project to spend its funds in the most effective way.

Data that are not used always cost too much. If the project managers and other potential users do not use or even read the reports of the unit, the entire cost of the monitoring has been wasted. If the users are involved right from the beginning in delimiting the work of the unit, they are more likely to use the results, especially if the results are presented in a format defined by the users themselves.

If too much emphasis has been placed on data gathering and the staff is unable to process and analyze the vast amount collected, the unit will be unable to present any results, which again wastes the entire cost of monitoring. Careful planning of data gathering and processing, with data gathering restricted to data really needed and processing starting immediately after the data come in will avoid such waste.

When choosing methodologies for data gathering and processing, one must also keep the costs in mind. If the amount of data is so large that computer analysis is needed, the cost of data processing will be greatly increased. One must ask oneself whether the funds allocated can cover the cost of expensive computer time and its programmer.

In many countries, transportation is likely to be a major cost. If the budget for transportation is small, the program should be based on enumerators and supervisors living in the villages and using bicycles for their work.

Material constraints should be identified before the program is designed, so that the project management can decide whether to allot more funds to the unit or whether they can accept a restricted scope of work.

When seeking a balance between reasonable objectives and sufficient resources, one should be conservative and realistic. Things are likely to go wrong more often than is expected.

4.2 Collecting information

Information can be obtained in various ways, which will be reviewed below. Most monitoring and evaluation units will use a combination of these.

4.2.1 Existing data

The first step in collecting information is to look for data already available. This search for existing data should not be limited to project documentation. Information may be available from a variety of sources, both governmental and private. In most countries, much research has been done at the farm level, and efforts should be made to obtain copies of the results. For many years now, experimental stations have been conducting trials on farming practices and have been collecting meteorological data; their reports may provide useful information on both the project area and the farming practices it proposes to promote. Results achieved in projects with similar goals can also be useful.

It will not be easy to get hold of all the desired documents, especially if they have to be obtained from other countries. Nevertheless a thorough search for existing data will save time in the long run because it will ensure that no time is wasted gathering data that are already available.
Some data required by the unit may be available from other sections of the project: Extension, Rural Credit, or Agronomy.
(More suggestions on where to look for existing data can be found in Chapter 16.)

Collecting existing data

4.2.2 Observations

A reliable way to obtain information is to go and observe the situation. But this approach cannot be used for all data, because it would require an enormous staff of enumerators.

For example: Suppose that a survey to measure income from petty trade were to be done by observations, it would require one enumerator for each trader, with the enumerator sitting in the shop all day and recording all the transactions being performed.

For some data, however, it can be very useful to have the enumerator observe the situation, which may be, say, the condition of a field or how a farmer handles a new piece of equipment. One enumerator can do many such observations. Observations are also useful in cross-checking data obtained in another way.
Many of the observations required in a farm study consist of measuring

a quantity (e.g. weighing the harvest of a sample plot, measuring rainfall, or measuring a field). Funds must be allocated for the equipment required for such measurements.

4.2.3 Interviews

As observations are expensive and are not appropriate for all data, most of the data are obtained by interviewing the farmer and the members of his household. One then has to rely on the information they are willing to provide. This information can often be checked, but one can never be absolutely sure that it is correct - not necessarily because a farmer does not want to tell the truth, but because he does not remember the exact facts or mixes up different events. Sometimes the farmer may try to please the enumerator by giving him the answer he seems to expect.

Data gathered by interviewing therefore will always include more possibility of errors than data gathered by observations. To reduce errors, the enumerator should do everything he can to establish a good relationship with the farmer and the members of the household. They are then more likely to tell him the truth and to make an effort to remember events. The way a question is asked can influence the farmer in his answer, or may help him remember the event correctly, so questions should be formulated with great care.
The interview approach can be applied in various ways:

- A few questions can be asked of all the people in an area. (This is how a census is done in most countries, to find out the number and the age, sex, and location of the population of a country.) Such interviews can be repeated at regular intervals;
- A few questions can be asked of some of the farmers in an area. These interviews can be a once-only event or be repeated as needed (once a year for some data; every few days for data on frequent activities);
- A very detailed study can be made of a few households. The interviewer then tries to find out all possible details on one or more of the activities of the households. This approach usually requires several interviews with each member of each household.

4.2.4 Trials

Trials at farm level are an effective way of obtaining data on innovations. In this way, farmers are asked to do certain things (grow a new variety, apply a certain quantity of fertilizer, etc.).

> For example: If a project needs to find out the effect of a high level of fertilizer on sorghum under existing farm conditions, it cannot wait until a farmer is found who uses that particular quantity of fertilizer on his sorghum. Instead, it selects a number of farmers who are willing to participate in the experiment and provides them with the necessary quantity of fertilizer to apply. The unit then gathers data on the crop results of these farmers.

4.3 Processing requirements

As already stated, care should be taken to allocate enough of the unit's resources to data procesing. Also important is to decide from the start whether the data will be analyzed with calculators or with a computer or micro-computer, since this has repercussions on the budget and on the qualifications of the personnel required.
A computer can perform calculations faster than a calculator and has a larger handling capacity, but these advantages become real only after the data have been stored in the computer memory in a correct format. This means not only having checked the data - a step that must be taken regardless of the analytical tool used - and entering them on recapitulative forms, but also coding, keypunching, and checking the stored data. In addition, existing computer programs will have to be adapted or special programs will have to be written to meet the needs of the unit. This can only be done by an experienced computer programmer and in the best of cases will require weeks - or more likely months - of work.

With a micro-computer it is possible to perform the less complex analyses more quickly than with a regular computer, provided that standard programs can be used and that the forms have been designed for that purpose. The data can then be ready for analysis sooner because the entire work is done within the unit.

Manual analyses can be handled by the unit's data processors right in the office, with no coding involved. In this way, the data and results are continuously being checked by people familiar with the surveys, to see whether the results are plausible. This is an advantage that no computer programmer can provide. The calculations still have to be verified, however, a process that can be time-consuming if a large body of data is involved.

When designing the survey program, one should make a rough estimate of the total number of data that will be collected and then calculate the time required to process and analyze them, either by hand or by computer. In most projects, only a few complex statistical calculations will be performed, usually on a series of rather small sub-samples. This is well within the capacity of a few data processors equipped with good calculators.

4.4 Reporting schedule

A decisive factor in the success of a monitoring system is the timeliness of reporting. Plans should be included in the program to make the results available to the users in a format appropriate to their needs and at a time when any desirable changes that the results may indicate can still be incorporated into the implementation effort. On the other hand, some data will become truly significant only when put into the perspective of an entire crop cycle. A possible schedule would be a two-stage system, with, say, a quarterly information sheet presenting preliminary results, and a full analysis at the end of each crop season, possibly including an analysis of trends over the years.

> For example: Shortly after the beginning of the season, information on the areas planted with each crop will be most useful if one wants to estimate the potential need for fertilizer. Early reporting on the use of fertilizer and insecticides can reveal whether extension efforts need to be reinforced. In contrast, an analysis of farm income or of the relationship between agricultural practices and yield can only be done at the end of the crop season.

5 GENERAL RULES FOR INTERVIEWS

5.1 Assuring the cooperation of the farmer

Successful interviewing requires cooperation on the part of the farmer. If he is to provide correct information, he should understand the question, remember the facts it refers too, and be willing to tell the truth.

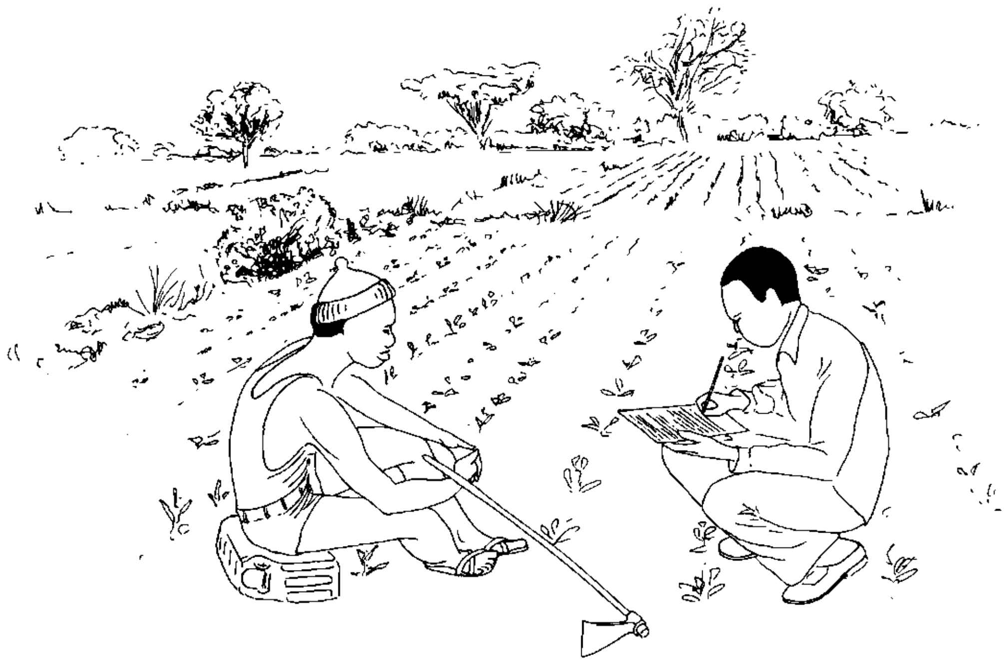
To encourage the farmer to make the effort to remember facts and answer truthfully, one should make sure that he understands why the study is being conducted. It is also important that farmers are not annoyed by interviews. This means that the enumerator should always be polite and pleasant, and should take the time to greet the members of the household according to local custom each time he comes to visit.

The time and place of interviews should be convenient for the farmers. Some do not mind interrupting their work in the field and may even welcome the break. Others resent the loss of time and prefer to talk at their house after the day's work is done. The time preferred for interviews may vary during the year depending upon the amount of work to be done in the field. The enumerator should be alert to the mood of the farmer. If his visit seems inconvenient, he should suggest coming back at another time.

When deciding where the interview will take place, privacy can be a consideration. There are things that a farmer will say to the enumerator, but not in the presence of a neighbour or even of another member of the household.

An interview can be tiring for the farmer so it should not last too long. It will be 'too long' if the farmer becomes tired and gives any answer just to get it over with, or simply refuses to go on. How long is 'too long' depends on the persons involved and on the difficulty of the questions. If interviews are to be repeated at regular intervals, it is especially important to keep them short (15-20 minutes maximum).

After the first few times, the farmers will have lost their curiosity about the enumerator and about the interviewing process, so they will become bored more quickly. Telling them the results of the survey as these become available and how that information is being used by the project can keep their interest alive.



Interviews are held wherever convenient for the farmer

5.2 Questionnaires

Much can be learned during an informal talk with a farmer but, if the interviewer does not have at least a list of the items that should be discussed, there is a risk that some may be forgotten. Preferably, interviews should be conducted on the basis of a questionnaire so that all enumerators ask the same questions. This allows the answers to be compared.

It is difficult to write good questions. Ideally, they should be short, precise, easy to understand, and be understood in the same way by everyone; they should not embarrass or annoy the farmer, or influence his answer. The questions should be checked by people who are familiar with the area to see whether they meet these requirements.

The questions should be listed in a way logical to the farmer. This might mean asking some questions which are really not essential for the survey but may help to put the farmer at ease or lead up to a necessary question.

Some surveys are done by interviewing farmers once only. For such surveys it is best to have each question fully written in the language that will be used in the interview. The enumerators read the questions exactly as they are written, and in the order in which they are written.

5.2.1 Type of questions

A question can be 'open-ended', which means that some space is left beside the question so that the enumerator can write down the farmer's answer exactly as he says it. This can be a fairly time-consuming process, especially if the farmer gives a long answer. Some enumerators then tend to write down only the most obvious part of the answer without trying to record more. With open-ended questions, it is difficult to obtain all the information at the same degree of detail.

An alternative is a 'multiple-choice' question, which has a list of possible answers already written down. The enumerator then encircles or marks with a cross the answer chosen by the farmer. Since one can rarely foresee all possible answers, some space should be left to record unexpected answers.

The advantages of multiple-choice questions are:

- They make it easy to record the farmer's answers without having to interrupt him to write them down;
- The answers are easy to read, check, and tabulate;

- The proposed answers can provide a memory aid to the farmer;
- All answers are at the same degree of detail.

A disadvantage of multiple-choice questions is that one has to foresee most of the possible answers in advance.

The type of question chosen - open-ended or multiple-choice - depends upon the topic and the objective of the survey. For a reconnaissance survey (first survey in an unknown region) in which the range of likely answers is not known, open-ended questions are appropriate and can bring to light important and unforeseen elements, even though the answers will be more tedious and difficult to analyze. In a more intensive survey done at a later stage, when most of the possible answers have been identified, multiple-choice questions are appropriate. The possible answers should always include an answer 'Others', so that any answer not listed can be recorded.

Example of a question worded (a) as an open-ended question and (b) as a multiple-choice question

(a) If you grew cereals last cropping season, what cereals did you grow? ....
....
....
....

(b) Did you grow cereals last cropping season? Yes/No*.
If yes, which cereals did you grow*?
1. White sorghum
2. Red sorghum
3. Millet
4. Maize
5. Others (Specify) ....

* Encircle the right answer. (More than one answer is possible.)

5.2.2 Wording of questions

The wording of a question is very important. An incorrectly worded question can be understood in different ways by the farmers, or they can give it a meaning different from that intended. Questions should therefore first be discussed with people familiar with the area and then tested on a few farmers.
Questions can easily be written in a way that influences the farmer's answer.

For example: If he is asked 'Did you use the improved variety of seed?', it is likely that his answer will be 'Yes' even if he did not because he knows that the extension agent is recommending it. If he is asked 'What variety of sorghum did you plant?', there is a better chance that he will indicate the variety he really planted.

The questions must be very precise and clear so that everyone understands them in the same way. A long, complex question will puzzle the farmer, who is not quite sure what is really being asked. It is better to ask several short questions than one long one.

For example: In answer to the question 'Last year, did you produce enough cereals so that you could sell some at the market?' the farmer might say 'No', meaning 'No, I did not produce enough to spare any for sale', or 'No, I did not sell any, although I had plenty of cereals', or 'No, I did not sell any at the market; a merchant came to my compound to buy'.

This question, worded in another way, is more likely to produce the correct information.

For example: Last harvest, did you produce more cereals than you needed to feed your household until the next harvest?
Have you sold any cereals since last harvest?
If so, where?

5.2.3 Problem questions

Information that seems very basic and obvious can still be difficult to obtain. Problems can arise in language because the staff of the monitoring unit may overestimate the farmers' knowledge of foreign terms. The staff themselves may use words of foreign origin in their local languages, but these words may not be known to the farmers. Another difficulty is that a word like 'family' can have various meanings: it may mean all the people with whom the farmer has ties of blood or marriage, or it may mean his wives and children only.

It is wise to avoid asking questions in more detail than is necessary. Whenever a question is written, the person writing it should ask himself why he needs to know the answer, and how precise this answer really needs to be.

For example: When a person is preparing a form for a household survey, he wants to know whether the individual members of each household are adults (and therefore workers) or children (who eat from the household pot but do not work). So on the form he writes the variable 'Date of birth'. In the rural areas of many African countries, that will create endless problems because few of the people are likely to know even the exact year they were born in. Unless a detailed demographic analysis is desired, it is better to simply ask the approximate age of each person, or to group ages in classes as a function of the work done in the fields.

Questions about intentions are often unreliable, as 'good intentions' are likely to prevail. Every farmer will say that he intends to follow the advice of the extension agent. One way to learn more about the farmer's real priorities is to ask a hypothetical question about a 'dream situation'.

For example: In a country where a national lottery is ubiquitous, one could ask: 'Suppose you won a million in the lottery, what would you do with the money?' This is such a farfetched possibility that people might feel free to speak the truth.

Of course, hypothetical questions only provide useful information in some cases, but they can be worth trying.
Sensitive questions should be avoided; the farmer is likely to answer them incorrectly and may remain annoyed and suspicious when answering the rest of the questions. What is considered sensitive is a matter of cultural definition, and one would do well to inquire about that among the field personnel. Questions about wealth (cash income, livestock owned) can be very sensitive, especially if the farmer is worried about tax obligations. Questions about food intake and grain stored can be upsetting to farmers who do not have enough food for their households.

5.3 Tables

When interviews are to be repeated regularly over a long period of time, a list of questions is likely to become tedious for both the enumerator and the farmer. In such circumstances, the questionnaire can take the form of a table, which can include columns for most of the likely answers. The enumerator then only has to fill in some numbers, adding any remarks that may be needed. This allows him to have a more relaxed 'conversation' with the farmer while filling in the table. The table itself can serve as a memory aid to the enumerator.
If the enumerator's notes on his observations in the field are entered in the same table as the farmers' answers during the interview, these two sets of information can be compared.
Tables are easy to check for completeness and accuracy, especially if the same sheet is used over several visits. The only drawback of tables in that during the first interview, the enumerator will have to take the time to explain to the farmers why each question is asked and what it means.

5.4 Frequency of interviews and observations

The number and frequency of interviews and observations required to gather data depend on the type of survey being conducted and the type of information being gathered. Basically, there are two types of surveys, both of which are appropriate in certain circumstances.

5.4.1 Single-visit survey

A single-visit survey is concerned with a situation at a certain point in time. The enumerator will interview or observe the farmer only once to obtain the desired information. Data on events that took place in the past (e.g. last year's crop yield) or on characteristics of the household and farm that do not change (e.g. place of birth of the farmer, location of the fields of the farm) can usually be gathered in one visit. The survey can combine interviewing the farmer (past events, family characteristics) with direct observations (location of the fields). A single-visit survey can be conducted over a larger area than a repeated survey because the enumerators can move on from village to village.

5.4.2 Repeated-visit survey

Within a monitoring unit, repeated visits are likely to be more usual than single visits. The frequency with which the visits are made can range from three or four times a year to every few days, depending on the type of information being collected. During the visits, the enumerator inquires about events that have occurred since his last visit (e.g. about agricultural activities or about income).
If the farmers being surveyed (or perhaps their children) know how to write, or even just how to write numbers, they might be able to note down some of the information required by the enumerator, who can thus visit them less often. In the beginning, this system requires much more care on the part of the enumerator and it is not always feasible.

How frequent the interviews should be depends not only upon the type of data being gathered but also upon the desired level of accuracy. Some events are clearly remembered for a long time.

For example: A farmer is likely to remember for months how much he paid for the ox he bought, because this is a rare event involving a large amount of money.

Surveys designed to gather data about such subjects may require only occasional visits (say once a month). Observations of some long-term processes (e.g. recording the performance of a crop) can also be done at infrequent intervals. In contrast, other matters require frequent observations.

For example: To survey the weeding of a field the field needs to be visited every few days so that progress can be observed.

Some events are quickly forgotten by the farmers, either because they are not considered important or because they happen so often that the farmer pays little attention to them.

For example: The farmer will know what he paid for an ox long ago, but may forget that he sold a chicken six weeks ago and what the price was, because he sells a chicken every few weeks.

5.5 Coding the answers

Coding means representing an item of information by a letter or a number which replaces it. The symbol used is not important as long as it is always used in the same way.

For example: In a survey, the letter M may stand for Male and the letter F for Female. It could also be decided to use 1 for Male and 2 for Female.

Coding is needed if the analyses are to be done by computer because only coded data can be handled by computers. It may be tempting to have the enumerator fill in the questionnaires directly in code to save space and make tabulations easier. Direct coding, however, should be introduced with great care. Coded forms are more difficult for the enumerator to fill in. They are also more difficult to check before tabulation. The unit director should consider whether coding is really necessary for the analysis and whether the educational level of the enumerator is high enough to make direct coding possible. If not, coding can always be done later by the data processors, after the questionnaires have been checked.

If coding is really essential for the analysis, codes close to the normal language can be used (e.g. the initial of the word they stand for). If that becomes too complicated, the possible answers, together with their code, can be printed on the questionnaire. The enumerator then encircles the code beside the right answer. It may take more paper to print such questionnaires but the results will be more accurate than if the enumerators write the code themselves. The same principle can be used in a table.

The following example recapitulates some of the suggestions made in this chapter.

Example of a questionnaire

In a questionnaire on the sale of livestock, one could simply ask: 'During the past week, did you or someone in your household sell any animals? If so, how many of each kind were sold?' and write down the answer. But it is possible that the farmer will only think about expensive animals, and not mention selling a chicken because it is such a routine occurrence.

The question can be made more systematic:

During the past week, did you or someone in your household sell any animals? If so, how many of each kind were sold?

11. Ox	Quantity ......
12. Donkey	......
13. Sheep	......
14. Goat	...1..
15. Poultry (chickens, guinea fowl)	...2..
16. Other animals: (Specify)	....

The enumerator circles the code number of the correct answer, which is also used as its code in keypunching for computer work, and writes down the quantity after the name of the animal. (In the example above, one goat and two chickens were sold.)
In a repeated survey, which is gathering more detailed information, the question is not presented in words but in the form of a table. During an informal conversation, the enumerator asks 'What type of animal was sold and at what price?' and writes down the number in the appropriate columns as shown in Table 5.1.

Table 5.1. Example of interview table for sale of livestock

Date of interview	Types of animals						Total price CFA
	Ox 11	Donkey 12	Sheep 13	Goat 14	Poultry 15	Others (Specify) 16	
10 May					2		900
10 May				1			7,000

When transferring the data to the computer, the code written in the heading of the column is used. There is no need to code the price, which can be copied directly.

6 PREPARING A SURVEY PROGRAM

In Chapter 3, it was shown how to delimit the work of the unit in consultation with the users of the information. In this chapter it will be shown how to prepare a work program on the basis of their requests.

6.1 General suggestions

When preparing a work program, the following points should be kept in mind:

- Right from the start, the program should cover both data gathering *and* data processing;
- The program should allow for an early start with data checking and data processing so that the work of the supervisors and the processors can be planned to start at the same time as the work of the enumerators;
- The forms on which the data will be gathered should, as far as possible, be designed for subsequent easy tabulation and analysis, although not at the risk that the validity of the data is reduced;
- The program should include a system of feedback from data processing to data gathering;
- It is easier to work from the general to the specific, first preparing a broad long-term plan of action covering several years and then proceeding to a detailed work plan for each crop season.

6.2 Long-term plan of action

A long-term plan of action need not be done in any great detail and should remain flexible. It indicates the types of surveys to be undertaken over a certain period of time, outlining their stages of development from reconnaissance surveys (the first rough gathering of information) to in-depth surveys.

Example of a long-term plan of action

Suppose that the monitoring unit is being asked to conduct surveys to satisfy the following requests:

- *Request A: Average yield and total production of all crops;*
- *Request B: Agricultural income per capita;*
- *Request C: Disposable income per capita.*

The results of the surveys will be used for the planning of agricultural extension and credit facilities, so fairly detailed information is required. The surveys will therefore have to be planned over a long-term period, say, three years.

In the first year, reconnaissance surveys will be conducted for all three requests. These will be single-visit surveys done with open-ended questions and performed on a large sample of the population. There may be only one sample, or there may be three different samples, one for each category of data.

In the second year, the results of the reconnaissance surveys will be used to design repeated surveys. Questionnaires will be revised to make them better adapted to the farmers' conditions and will include elements that were overlooked in the first year. Some of the questions asked in the first year will not be asked again, because they appeared not to be relevant, or the results were the same for every farmer, or the information is now complete. Improving surveys in this way will be possible if enumerators are encouraged to note on the survey forms any relevant additional information obtained, either through their own observations or through their contacts with the farmers and the village people.

In the third year, the results of the second-year surveys will be used to design in-depth surveys on small samples of the population.

The long-term plan of action for Request C is as follows. In the first year, data will be gathered to identify the main income-earning activities. In the second year, when the farmers have become used to the survey, data gathering will take less time so the questions can be more detailed, especially for the more important activities. In the third year, the sample can be split into different sub-groups and efforts will be concentrated on collecting detailed information on specific topics such as the income distribution within the family or the consumption pattern.

6.3 Program for the first crop season

Unlike the long-term plan of action, the program for a crop season must be designed in great detail, although it should be flexible enough to handle unforeseen events. The program defines the type of surveys to be done, the questionnaires to be used, the samples to be selected, the frequency of the surveys, and the organization of feedback and validity control.
The surveys to be done in the first crop season are all reconnaissance surveys so do not require very precise data. A margin of error of about 30 per cent, usual in such surveys, is acceptable.

Example of a program for a crop season

The first step in preparing the program is to make a list of the data required to satisfy each request. The data will be gathered on samples of the population. For Requests A, B, and C of Section 6.2, the lists are as follows:

Request A: Average yield and total production of all crops
- Area under each crop;
- Production figures for each crop;

- Average yield per crop;
- Total number of farming households in project area.

Request B: Agricultural income per capita
- Yield figures for each crop;
- Direct production costs for each crop (seed, fertilizer etc.);
- Indirect production costs (tools, draught animals etc.);
- Size of the household;
- Farm area;
- Area under each crop per farm;
- Market prices of crops.

Request C: Disposable income per capita
- Same data as Request B;
- Income from trade, livestock, other sources;
- Cost of living (food, taxes, other essential expenditures).

Looking at these lists, one can see that some of the requests require the same data (e.g. crop production figures, size of household). This makes it possible to economize on data gathering by combining surveys.

No requests have been made for an analysis of the interactions between survey results (e.g. between agricultural income and disposable income). This allows the unit to use different samples of the population to gather data. Data gathered from different samples are likely to be more valid than those from a single sample: more farmers are involved, so the frequency of interviewing will be less and the farmers, being interrupted less often and for brief interviews only, are likely to be more cooperative. Because the samples will all be chosen from the same population, the results of the surveys can be aggregated and used together, provided that the samples have all been chosen in the same way (e.g. at random).

A possible set of surveys to satisfy Requests A, B, and C could be the following:

Survey 1: For all the fields of a sample of farmers, yields (expressed in kilograms per hectare) will be estimated (by sample plot). The total production will be surveyed so the area under cultivation can be calculated. The direct costs per field and the indirect costs per farm will be surveyed on the same sample. Most of this work will be concentrated at the end of the crop season.

Survey 2: The number of persons living in the households will be counted and their cost of living estimated. This survey can be done at any time during the crop season.

Survey 3: Market prices will be collected at local markets at regular intervals (about once every two weeks).

Survey 4: Income from livestock will be obtained in weekly or fortnightly surveys.

Survey 5: Income from trade will be obtained in weekly surveys.

Survey 6: Income from other sources will be obtained in weekly surveys.

The results of Survey 1 provide the data required for Request A. The data on production of the field and the production of a sample plot provide data on the area under each crop and the average farm size. To estimate the total production in the project area, the total number of farming households is required; this can be obtained from the project management.
To satisfy Request B, the data for Request A are supplemented by the results of Surveys 2 and 3.
Request C can be satisfied by using the data for Request B, supplemented by the results of Surveys 2, 4, 5, and 6.

Two things will now be clear:
- The data needed to satisfy one request need not necessarily be gathered in one survey;
- The data needed to satisfy different requests can be grouped together and obtained in one survey.

When preparing the program for a crop season, the unit's material constraints should, of course, be kept in mind.

6.3.1 The enumerator's work load

The work load of an enumerator has to be reasonable. It is not wise to plan a maximum load of interviews for every working day because there will always be unforeseen delays and absences, whether on the part of the farmer or on the part of the enumerator.

> For example: If an enumerator is responsible for interviewing sixty farmers, it is not wise to plan that he visit six farmers a day over a period of two weeks. It is more realistic to plan that the enumerator visits, say, only four farmers a day, which allows him time to catch up, if need be. Alternatively, it could be planned that he visits five farmers a day but for only four days of the week, allowing him one day a week to catch up. In either way, he will see each of the sixty farmers once every three weeks.

A local market

The dates for single-visit surveys should be planned for when the enumerator's normal work load is not too heavy. No additional surveys should be planned, for instance, while enumerators are engaged in crop estimates. Nor should extra surveys be timed to coincide with periods when the farmers are very busy or likely to be travelling.

If an additional survey is to be done - once only - on a sample already being regularly interviewed, it can be done during one of the regular interviews. Then, however, the enumerator should only be required to visit one or two farmers a day, because he will need time to explain the additional survey to them.
When an additional survey is to be done on a new sample, the work load should be even lighter because the enumerator may have to spend a lot of time introducing the new survey to the farmers: he has to call a meeting of the farmers to explain the survey, after which he has to visit each farmer in the new sample, explaining again what the survey is for. Only then can he perform the survey.

6.3.2 Feedback

Prompt feedback of results enables the users to take whatever action may be required while the project is being implemented. Feedback also enables the monitoring unit to modify its surveys and develop additional ones if necessary. The possibility of such changes should be left open in the original program. Excellent sources of feedback are the enumerators because they are aware of problems in data gathering and may have useful remarks about the survey results.

Some effects of feedback are illustrated below:

- After a period of data gathering, it becomes clear that the results obtained from one of the surveys are constantly wrong or incomplete because of difficulties with the farmers. Knowing this, the unit may try to re-word the questionnaire or devise another way of obtaining the data. Having tried that without success, it can then decide to stop this survey or to remove the problem questions;
- After a while, the results of a survey become so regular that they are predictable. The survey has fulfilled its purpose and can be stopped, even if it had been planned for a longer period;
- The results of the first analysis of a survey show that the unit had overlooked an important development in the villages when designing the program of data gathering. Together with the users of the information, the unit can then reorganize this part of the program and add

the relevant questions;

- The results of a survey cannot be fully analyzed or explained because the information is inadequate. Rapid preliminary analysis shows which information is missing and the survey can be modified, or a new survey can be started, to obtain the missing information.

  For example: The unit has been requested to analyze the relationship between fertilizer use and the yield of cereal crops. To do so, it has gathered data about fertilizer input, yield, and damage to the crop. When analyzing the first data sets, the unit found odd results, which could probably be explained by the timing of the fertilizer application. The supervisors were then instructed to have the enumerators collect this missing information. By then it was too late to obtain precise data, but the farmers were able to remember whether they had applied the fertilizer when planting or while weeding.

6.4 Choosing the groups to be surveyed

The population to be surveyed can be all the farms in the project, but it can also be the people, the households, the fields, animals, or objects in the project. Users will request information on a population, but this does not mean that they need data on each and every element of that population; what they want is a general idea.
It is tempting to gather information on every element of the population to be sure that the results are correct. This, however, would be a costly and time-consuming exercise, producing enormous amounts of data that would have to be processed and analyzed. Long delays would ensue before this information could be used.
Usually, therefore, those requesting information prefer timely reporting of results, even if it means less precise data. For this reason, surveys on the entire population are very rare. Instead, the information is gathered on only a sample of the population, which is selected in such a way that it can be considered representative of that population. A survey done on the sample gives roughly the same results as a survey done on the population.

6.4.1 Survey on a sample

To be sure that the sample represents the population, it should be selected in accordance with certain rules. It is never possible to select a sample which is representative of the population in all its aspects, but it is possible to chose a sample which is representative in those topics to be studied in the surveys. This is all that is needed. A representative sample means that the findings from that sample apply to the entire population.

For example: The farmers in a sample have produced an average of 1,800 kg of sorghum per farm. If the sample is representative of the population for the production of sorghum, one can conclude that the 2,000 farmers in the population have also produced an average of 1,800 kg of sorghum per farm. So the total amount of sorghum produced by the population is about $1{,}800 \times 2{,}000 = 3{,}600{,}000$ kg.

The procedure used to select a sample depends on how much is known about the population, on whether it is highly diversified or not, and on the way in which the data will be analyzed. (Methods of selecting a sample are described in Chapter 17.)
While respecting the rules of sampling as far as possible, one should keep some practical items in mind. If the population is spread out over a large area and transportation is difficult, it is wise to survey clusters of farmers living in the same neighbourhood. It is unwise to place an enumerator in a village which is difficult to reach during the rainy season, as he will then have to manage on his own for long periods without a visit from his supervisor.

6.5 Designing survey forms

6.5.1 Questionnaires and tables

The first step in drafting a survey form is to list all the subjects on which information is needed, in as much detail as possible. The subjects are then grouped into categories. Within each category, the subjects are placed in the most logical order, both for the farmer and for the enumerator. This logical order may differ from that used during analysis. Whenever possible, the variables for similar topics should be listed in the same order. In a repeated survey, it is especially important to establish some clear routine, so that the farmer knows what he will be asked; he is then more likely to remember the information. (This is especially true of financial transactions.) Data are also easier to tabulate if they are always in the same order.

For repeated surveys, a new form can be used for each interview, or the same form can be used for several interviews. If a form is used only once, it is easier to design, but a form used for several interviews has other advantages: it uses less paper, makes processing easier, cuts down on editing, and facilitates validity control. This is especially true if tables are used rather than questions.
A tentative text can then be written for each question or each table, in which cross-checks should be included to help in checking the validity of the data gathered. A cross-check means asking for the same information twice, but in different ways. As was explained in Chapter 5, the wording of the questions is very important, so the texts should not be written in a hurry. It is best to prepare a draft, read it again after a few days, and pass it on to several people for comments, especially to enumerators.

Sometimes the person designing the questionnaires or tables does not know the local language in which they will be used, so he writes the first drafts in the national language. He should then discuss each question with several speakers of the local language and have them agree on its translation. Key terms such as family, farm, field, or ownership of land can be translated in different ways, so the intended meaning of the term should be thoroughly discussed. The accuracy of the translation can be checked by having someone else translate it back into the original language. The final draft survey form should be written in the local language.

The full set of survey forms should be checked for completeness against the original list of information requested. If some of the required information has not yet been included, the questions or tables should be

revised accordingly.
The layout of the forms (location and space for each question) has to be decided. Normally, the same layout should be kept for the entire duration of the survey, otherwise the enumerators may have trouble in finding the right column, which will reduce the validity of the data. It is prudent to check whether the layout is practical after a few pages have been prepared. It should also be remembered that the questionnaires have to be filled in according to a strict time schedule. This schedule cannot always be kept, so the forms should make allowance for this. The enumerator should be able to do the survey off-schedule and still record the results on the same form. If that is not possible, special forms will have to be prepared to enable enumerators to catch up when a delay has occurred.

Once a questionnaire has been agreed upon by the people designing it, a draft should be typed and discussed with the field supervisors and enumerators. At this stage, several copies of the questionnaire should be filled in with data to find out whether the columns are of the right width, whether the lines are comfortably spaced, and whether the headings are easy to read and in the right order. Sometimes abbreviations will have to be used because of lack of space. Ensure that such abbreviations cannot be misinterpreted.

Now the draft is ready to be tested under working conditions, during interviews with a few farmers. This will reveal whether the questions and tables are workable, and whether the layout is practical for use in the field.

Example of a survey form for agricultural activities

The unit has to collect the following data, as agreed upon in discussions with the users:
- Crop grown, area, production, and yield;
- Timing of agricultural activities;
- Agricultural techniques used;
- Use of fertilizer.

Table 6.1. Survey form for agricultural activities

Crop season: .... Crop: ....

Name of farmer: .... Field: ....

CALENDAR OF ACTIVITIES

Type of work	Period		Technique used		Observations
	Date begin	Date end	Manual	Animal traction	
Land preparation					
Ploughing					
Sowing 1					*
Sowing 2					*
Sowing 3					*
Fertilizer appl.					*
Weeding 1					**
Weeding 2					**
Weeding 3					**
Thinning					**
Ridging					**
Harvest					

* Note variety (type) and quantity in farmer's unit of measurement

** Note whether completed

Has the crop been damaged? Yes/No

If so, indicate the type of damage: drought, flood, insects, cattle, wild animals, fire, other: ....

OBSERVATIONS

....

....

....

Table 6.1 (contd.)

HARVEST

Sample area: .... ha

Harvest of sample area: ..... kg on head: ..... kg after threshing

Total production of field in number of baskets harvested: .....

Weight of three baskets chosen at random

Weight of first filling: .... kg

Weight of second filling: .... kg

Weight of third filling: .... kg = after threshing .... kg

Weight of average filling: .... kg

OBSERVATIONS

....

....

....

CALCULATIONS

Total production on head:	.... × .... = ...	kg
% of weight left after threshing:	.... ÷ .... = ...	%
Total production in grain:	.... × .... = ...	kg
Yield estimate sample area:	.... ÷ .... = ...	kg/ha
Area of field harvested:	.....	ha

The data will be used to describe the existing farming system and to develop an extension program. The information will be recorded as the crop season progresses. Some data (timing of agricultural activities) will be gathered by frequent, repeated interviews, but others (crops grown, fertilizer used) will be obtained in one interview at the proper time. An easy way to record the data is on a set of forms, one for each field, which the enumerator keeps up to date throughout the crop season. Table 6.1 shows the two sides of such a form, assuming one crop per field. It includes space for all the data and for basic calculations to be made from those data. For mixed cropping, the same basic form can be used, expanded a little so that the crops can be recorded individually.

Data on events and sources of income that occur regularly throughout the year can be obtained through questionnaires, but can far better be obtained in a set of tables, to be updated regularly. This system is appropriate for data on livestock production, crafts, trade, and so on. Such tables can cover several interviews and thus save time and paper and make tabulation easier.

Example of a survey form for repeated interviews

A farmer's income from crop cultivation is being monitored through a set of surveys. One of them is a repeated survey on all the expenses he incurs with small agricultural tools and the places where he purchases these tools or has them repaired. It is decided that this survey should be conducted every two weeks, because the farmer is not likely to remember small repair expenses for much longer. As an open-ended questionnaire to be used once only, it could be designed as follows:

Name of farmer: .... Village: ....
Date of inquiry: ....
Did you purchase any agricultural tools (small hoe, machete etc.) since the last inquiry? ....
If so, indicate for each tool:

Tool:	Price:	Where purchased: ....
....	....	....
....	....	....
....	....	....

Did you have any of your agricultural tools repaired since the last inquiry? .....
If so indicate for each tool:

Tool:	Cost of repair:	Where repaired: ....
....	....	....
....	....	....
....	....	....

This calls for a lot of writing on the part of the enumerator and the results are difficult to process, because each inquiry is done on a separate form.
A multiple-choice questionnaire per interview would be more practical for the enumerator but still tedious to tabulate. The most practical way of all is to use a table such as Table 6.2. Here the enumerator is free to word the questions in whatever way he likes as long as he covers all the subjects. Before such a table is used, it should be explained thoroughly to the enumerator to avoid errors on his part.

At the start of each interview, the enumerator notes the date of the interview on the first open line of the table. He asks the farmer whether he has bought any tools or had any repaired since the last interview. If so, he asks how much the farmer paid for it. He writes the number of each type of tool and the corresponding price in the appropriate columns, using one line for each transaction, and marks an X in the column corresponding to the place where the transaction occurred. As can be seen in Table 6.2, the farmer informed the enumerator on 10 May that he had spent 500 CFA to buy two small hoes at the local market and 100 CFA to have his hatchet repaired by the

Table 6.2. Purchase and repair of hand tools

Name of farmer: .... Village: ....

Date inter-view	Number purchased or repaired						Total price (CFA)		Location of action				Observations
	Small hoe (1)	Large hoe (2)	Machete (3)	Hachet (4)	Rope (5)	Other (6)	If purchased	If repaired	Local market (1)	Town market (2)	Smith (3)	Other (4)	
10 May	2						500		x				
10 May				1				100			x		
24 May													Nothing
TOTAL													

smith. On 24 May he had had no new expenses. At the end of the period the total expenses for purchase and repair can be calculated right on the questionnaire. The tabulation is kept up to date in the office. The numbers in each heading are codes for keypunching the data.

Another type of form - a recapitulative form - should be designed at the same time. Recapitulative forms are forms on which data already noted on various survey forms are regrouped. These forms are very useful for checking the validity of data and allow an early start to be made with tabulations for rapid reporting. A first draft of the recapitulative forms should be prepared when drafting the data-gathering forms, so that they can all be tested at the same time. (Details on format and use of recapitulative forms will be found in Chapter 7.)

6.5.2 Manuals

A manual describing the work program of the enumerators and informing them how to fill in each survey form should be prepared as the program is being organized. The enumerators should have the manual with them when they begin the formal surveys. Even if the supervisors visit the enumerators frequently, the enumerators will be alone during most of their work and need a clear, detailed reference manual.
The manual does not replace the explanation which is given about the program of surveys and the use of the forms during a training session at the beginning of the surveys (Chapter 14). The manual is a summary of these explanations, with the following key elements discussed in detail:

- The objectives of the survey program, and the purpose of each survey;
- The deadline for data gathering for each survey, with an explanation of the reason for the deadline;
- Methods of performing the necessary measurements and for filling in the forms;
- How to use and maintain the equipment;
- The frequency of data gathering for each survey;
- What to do when an interview cannot be done on time.

A manual can never be complete because not all problems that might arise during surveys can be foreseen and because other surveys may be added at a later stage. For these reasons it can be practical to place the manual in a loose-leaf binder so that supplements can be added later. The importance of using the manual should be pointed out well before the start of the surveys, at which stage the enumerators should learn where the explanations for each survey can be found and how to use the manual when they are on their own.

Preparing such a manual is a tedious job, but it is an important tool in preventing errors in data gathering. It is also very useful to the unit director, because it forces him to state clearly what he wants, how he wants it, and why. It is likely that he will revise some forms when writing the manual.

An introductory book for enumerators (such as the one accompanying this book, or the publications of the *Institut Africain pour le Développement Economique et Social*, INADES) provides useful background information but cannot take the place of a manual written specially for a survey program.

Each enumerator should also be given a precise time schedule of his work for the first period of his inquiries when he has not yet developed a routine for his rounds of interviews and field visits. The supervisors should also receive detailed instructions on how often to visit each enumerator. During the first few weeks of data gathering, the enumerators should be visited frequently by their supervisor and possibly also by the unit director. Even in the best planned surveys, enumerators are going to face many unexpected problems. It is important to show them how to solve such problems while these are still fresh in their memories.

The supervisors and data processors should have the same manual and possibly also another manual describing their role in tabulation and analysis and outlining their work schedule. This manual should discuss the use of each form and the procedures to be followed in checking the data.

The purpose of these manuals is to provide guidelines on how to perform the job correctly and what to do when a problem arises. They also help to bridge the gap between the workers in the field and those in the office.

6.6 Introducing the surveys to the farmers

Early in the process of designing the surveys, the unit director should contact the traditional chiefs of the villages to explain the surveys to them and to ask for their cooperation. He should make similar contacts with the administrative authorities. Both these groups of people can provide him with information useful for the surveys.

When the time comes to commence data gathering, the director should ask the village chief to call a meeting of the villagers. At this meeting, the chief will introduce the director, the supervisor, and the enumerator who will work in the village. The director will then explain to the villagers the reasons for the surveys. If the director does not speak the local language, this explanation can be given by the enumerator.

The traditional chief welcomes the director to the village meeting

It is important to take the time to explain the surveys properly to the farmers as this will greatly influence the extent to which they will be prepared to cooperate later.

If possible, the sample of farmers who will take part in the surveys should be selected at that meeting. This will show them that the selection is entirely at random.

It may be decided that the enumerator will call another meeting of the sample farmers to make more detailed explanations or, alternatively, he will make appointments for the first interviews. When the enumerator visits each farmer for the first time, he will have to repeat the explanation of the surveys to make sure the farmer has understood. The farmer may be more willing to ask questions in the privacy of his own home than at a meeting.

Often the farmers are greatly interested in learning about their production and income. If they know they will be given this information at the end of the crop season, this can be a strong incentive for them to cooperate.

6.7 Testing the forms

When the final drafts of the questionnaires have been compiled and the work programs of the enumerators and supervisors have been drawn up, the questionnaires and work programs should be tested in the field on a few farmers. Ideally the farmers chosen for the test survey should not be the same farmers who will be interviewed during the real surveys, as it would only confuse them. In practice, however, this can be difficult to arrange.

Testing the forms in the field will reveal whether anything has been forgotten, whether questions are being misinterpreted, and whether answers are being obtained that cannot be recorded properly. The unit director, the supervisors, and the enumerators should all participate in this testing.

When conducting test interviews, the enumerator should observe the farmers' reactions to the questions and not just write down their answers. He should note which questions have to be re-worded before the farmer can answer, which are not answered at all, which lead to long discussions before agreement is reached on an answer, and which are not clear in the local language.

The unit director may have to insist on a testing period, because project managers, not familiar with survey procedures, may think it a waste of time. It is worthwhile insisting, however, because testing greatly increases the validity of the data gathered later. It also helps to avoid disruptive changes during the crop season and therefore leads to better analyses. In addition, testing provides useful on-the-job training for everyone concerned.

The testing period for repeated surveys should be long enough to allow for repeated visits to the farmer with the frequency planned for in the surveys, to check whether the farmer does indeed recall the data required. For single-visit surveys, the testing will take only a couple of days.

The data gathered during testing should not be used for any real analysis, but should be tabulated and analyzed to check on the practicality of both the data-gathering forms and the recapitulative forms, and also to give the tabulators some practical training. The results of these test analyses may possibly reveal elements that should be investigated further in the surveys, or point out that some questions are superfluous because the answers are always the same.

When the testing has been completed, the results should be discussed with the entire staff. During this discussion, everyone should be encouraged to make suggestions on the format and content of the questionnaires, and also on the work schedule that will be required of the enumerators and data processors. There will be less occasion for them to complain later that their workload is too heavy if they have participated in its preparation.

After this testing, the program for the first crop season is ready. It covers:

- The surveys to be made, with the exact questionnaires and time schedule for each;
- A list of data to be obtained from sources other than surveys;
- A precise work schedule for the enumerator (interviews, observations in the fields, paper work, and training);
- A flexible program for the supervisor;
- A program for filling in the recapitulative forms;
- A schedule for coding and keypunching, if necessary;
- The work program and instructions written in a manual so that each staff member knows precisely what he has to do.

7 THE VISIT AND TABULATION SYSTEM OF VALIDITY CONTROL

Processing data that are incorrect is a waste of time and money, so all possible care must be taken to ensure that the data are complete and accurate. The most effective way to do this is for the supervisor to check the data in the villages - with the enumerator - as soon as possible after the data have been gathered.

This chapter will describe the visit and tabulation system of validity control. The system combines frequent checks on the work of the enumerator, with on-going tabulations (on recapitulative forms) of his results as they become available. It requires that the unit director or a supervisor visits each enumerator frequently. Making frequent checks on the work of the enumerators in the villages is probably the most crucial step in guaranteeing the quality of the final analyses. This is why supervisors should always be included among the staff of large units. The director of such a unit could not possibly handle the double work load of frequent regular visits to enumerators in addition to his tasks of designing, coordinating, and managing the program. Nevertheless, the director should visit the survey sites fairly regularly to keep himself informed of the realities in the field and to encourage the enumerators and supervisors, even if the bulk of validity control is done by the supervisors.

The supervisor's visits to the enumerator have three related objectives:

- To check the validity of the data gathered, and verify that the enumerator is doing his work regularly and carefully;
- To provide assistance to the enumerator in case of difficulties and to explain the surveys further as needed;

- To update recapitulative forms and to identify any trend in the data as early as possible.

The first two objectives are fulfilled at the same time so they will be discussed together in the next section. The use of recapitulative forms will be explained in Section 7,2.

7.1 Checking and helping the enumerators

The enumerators are quite independent in their daily work and the supervisors have to trust them to a certain extent. Nevertheless, it is important to check the work of the enumerators by doing cross-checks on the completed survey forms and by verifying some information in the field. The data that can be checked depend upon the kind of surveys being done, so the following remarks on validity control are merely indicative.

Each supervisor should visit his enumerators at least once every two weeks unless they are very experienced, and even more often during harvest time or when the enumerators need help in extra tasks such as measuring fields or picketing off sample plots. The supervisor should usually come unannounced, to see whether the enumerator is actually working and whether he is keeping to his work schedule. It is convenient to give each enumerator a slate to hang on the door of his house; every morning he writes on the slate where he is going (name of farmers to be visited, location of fields, etc.) so that the supervisor will know where to find him.

7.1.1 Checking the work schedule

Since each enumerator has a precise work schedule for interviews and other activities, it is easy to check whether he is indeed working to schedule. Delays are always detrimental to the quality of surveys, so all efforts should be made to see that they are avoided. Unavoidable delays will occur if the farmer or the enumerator becomes ill or if the enumerator cannot go to the fields because of heavy rain, and so on. Enumerators will inevitably get behind schedule on occasions, which is why survey programs should include instructions on how to conduct delayed surveys. Such instructions should be included in the enumerator's manual, together with suggested time schedules for catching up if delays occur.

Each enumerator should also have a schedule, in writing, for such work as measuring fields, picketing off sample plots, and so on, so it is a simple matter to find out if he is working to schedule. If an enumerator is far behind schedule, the supervisor and he should work out a special schedule to catch up, in the light of the specific situation. If an enumerator has not kept his schedule and has no valid excuse, the supervisor should increase the frequency of his visits and come back soon afterwards to check that the schedule given to catch up is indeed being followed.

7.1.2 Checking the survey forms

The supervisor should check the survey forms to see whether they are completely and correctly filled in. This will be time-consuming but has to be done because it is only soon after the inquiry has been completed that one still has the chance to correct errors and rectify omissions. If an error or omission is discovered at the office, it is usually too late and the form will be rejected; the effort already expended will have been in vain.

There should be cross-checks built into the forms, so that information collected in different places or by different surveys should agree. These cross-checks should be at least partly explained to the enumerator so that he is able to check the data he obtains from the farmer. Cross-checks can be very simple.

> For example: Suppose a labour survey is done at the same time as a survey on the use of fertilizer. If the farmer reports during the labour survey that he has applied fertilizer, an entry about fertilizer should be found on the fertilizer survey form as well. Observations made by the enumerator in the field should correspond to the answers on the forms, and so on.

The supervisor, who will have made himself familiar with the farming practices in the area, should be able to recognize information that is unlikely, and should ask for an explanation, first from the enumerator and then, if necessary, from the farmer himself. This applies for all answers or combinations of answers outside the usual range. It should always be remembered that an item of information out of line with the rest needs some explanation, otherwise people at the office, who will also be checking the data, will think that the odd result is due to carelessness on the part of the enumerator.

If the supervisor finds that all the farmers' answers are very similar, he should ask the enumerator to explain this so as to be sure that the enumerator did not simply complete the forms in his own home without bothering to go and ask the farmers.

After the supervisor has checked the data, he should initial the form or denote in some other way what he has checked.

7.1.3 Checking observations

Data gathered directly by the enumerator (e.g. field measurements) should also be looked at critically by the supervisor. He should make some spot checks, choosing a few fields at random and roughly measuring them by pacing. The supervisor can also check the placement of sample plots.

Crop production is more difficult to check but a talk with some farmers who may be willing to show him their granaries can be a useful rough check. The supervisor should go and look at some of the fields before they are harvested. Making a tour of the fields is a good way of observing the performance of the crop. This knowledge can be used later in checking the results obtained by the enumerator. After the harvest, the residues of the crops on the fields can also help to indicate whether crops were very good or very bad.

If a market survey is part of the survey program, the supervisor should visit the markets once in a while to inquire about prices and check the weights of local measures.

Example of checking data

A low yield is calculated for a field that had been observed to be good or average. This should be checked first with the enumerator: he may know that the crop was damaged by cattle just before harvest (information he should have written on the form), or that the heads of grain were empty. If the enumerator is not able to supply a satisfactory explanation, the reason should be checked with the farmer. This checking can include measuring his granaries, if need be, or checking the area of the field by pacing. If the production has been calculated both before and after threshing, the percentage of loss by threshing can also help in spotting unlikely data. As the supervisor becomes more familiar with the farmers in the sample, he may be able to spot the reasons for unusual or unlikely data.

7.1.4 Using checks for additional training

The entire checking process will be more productive and more beneficial to the enumerator if the supervisor does it by working *with* him, explaining why a doubt arises about an item of information, rather than behaving in an authoritarian manner. An enumerator is likely to know quite a lot about the farmers; if the supervisor shows him how to use this knowledge and his common sense to do a first check himself, this will build a good relationship between them and will ensure better data. When going to the fields together, the supervisor can show the enumerator what he should be looking for and how to write the information down on the forms. The supervisor should also greet the farmers in the fields, showing them that the work of the enumerator is really considered important by his superiors.

Even in the best of circumstances, it is likely that some survey forms will be marked as incomplete or doubtful and unusable for analysis. This is always detrimental to the validity of the survey. If an enumerator consistently does poor work, the supervisor should have the authority to reprimand him or to recommend that sanctions be taken against him by the unit director.

Additional training

The supervisor should inquire whether the enumerator has had any difficulty in filling in the forms or in performing any other task in his program. Even after good training and with a well-prepared manual, the enumerator may have a useful item of information which he does not know how to record. He should be encouraged to make a note of this on the form, so that he and the supervisor can decide how to handle the matter. The supervisor can later discuss this with his colleagues and the unit director so they can issue the same instructions to all the enumerators.

If the enumerator has had difficulty in measuring a field or placing sample plots in an unusually shaped field, the supervisor should do that job with him, taking this opportunity to teach the enumerator how to handle that task in the future.
It sometimes happens that the enumerator has more work than he can handle. If so, the supervisor should arrange that an enumerator from a neighbouring village comes to help him for a couple of days, or that a boy from the village be paid to assist. Such assistance is only useful

with jobs like measuring fields and placing sample plots; interviews and evaluations of harvests should be done by the enumerator himself.

Even if no problems seem to have arisen, the supervisor should take the time to discuss any aspects of the survey that have given problems to other enumerators. During this discussion, he will often discover that the same aspects were not yet clear to this enumerator either. This offers yet another opportunity for training, which is a major task of a supervisor.

7.2 Recapitulative forms

On-going tabulations on recapitulative forms regroups data from the survey forms being used by the enumerators. Presenting the data on one subject on one recapitulative form, they are updated by the supervisor at intervals as more information becomes available. Besides offering a valuable way of checking data, recapitulative forms allow early preliminary reporting. Data that would normally not be received from the field until the end of the crop season are made available much earlier. Even if the information is still incomplete, presenting some results quickly to the users ensures rapid feedback from them.

Example of the use of recapitulative forms

Suppose a survey is being conducted for the extension service to find out which innovations are being accepted by the farmers and whether these are leading to an increase in agricultural income. The innovations are animal traction, mono-culture, improved seed varieties, and fertilizer. Throughout the crop season, data are being gathered on:

- Type of field preparation;
- Area cultivated;
- Type of seed used;
- Type and quantity of fertilizer used;
- Yields obtained.

The information is being recorded on survey forms, one sheet for each

field throughout the season. No information is being collected about labour time because labour is not considered a constraint. Nor is information being gathered on the quantity of seed used because this information is difficult to obtain with the means available to the unit. Yields are found from estimates made on sample plots of 10 m^2. Crop production is estimated by counting the number of times the harvest baskets are filled when the harvest is being transferred to the granary. A few baskets are selected at random and their contents weighed. The contents of one of those baskets is then threshed and weighed again. The area cultivated is roughly estimated by dividing the total production by the yield. Information about damage to the crop is included to explain unusual figures.
The results of all this data gathering are being summarized on recapitulative forms - one for each crop - which the supervisor prepares at the beginning of the crop season. An example is shown in Table 7.1. Each time the supervisor visits an enumerator, he updates the recapitulative forms with the information that has become available since his last visit.

The data entered on the recapitulative forms are taken direct from the original survey forms on which the enumerator is noting the data obtained from the farmer. If some calculations are required, a column should be reserved for this on the sheets (and also for any calculations to be made later at the office). The form should be designed in such a way that only simple calculations are required in the field; these should always be kept to a minimum.

For example: It may be necessary to convert the crop production figures of each farmer into yields (kg/ha) before entering them on the recapitulative forms. In this way, the achievements of the farmers can be compared. Once the yields have been listed, the extreme (high or low) yields can be spotted. These extremes can be very important because, if proved correct, they indicate both the risk and the potential of crop production under farming conditions.

It may also be necessary to convert some of the data to sums or averages before they are entered on the recapitulative forms.

Table 7.1. Recapitulative form per crop

Crop season: .... Dates updated .../.../.../.../.../.../...

Region: .... Crop: ....

Identification		Area	Seed	Fertilizer			Production			% of weight after thres-ing	Damage ?	Observations (explain type of damage, unusual household situation etc.)
Farmer	Field	harv-ested ha	Var-iety	Type	Quantity		Sample plot kg	Field				
					kg	kg/ha		kg	kg/ha			
Total												
Means/Field												
Means/Farm												

With recapitulative forms, the supervisor can see not only whether the enumerator is up to date with his work, but also whether there are any discrepancies in the data. If these discrepancies have not been clearly explained, they should be checked.

For example: If all the farmers in a village have sown some maize except one, the supervisor should make sure that the enumerator has not simply forgotten to inquire about maize with that farmer. If that farmer has indeed not sown any maize, the reason why should be ascertained, if possible, and noted on the recapitulative form.

In the same way, if a farmer has sown a crop but then has to abandon it, this should be noted as well.
If a farmer obtains a yield much smaller or much larger than that obtained by the other farmers, that too should be checked and, if found to be true, should be noted, with possible explanations.

The recapitulative forms are a great help in checking the validity of the data, because even the most careful supervisor will overlook some missing or incorrect data once in a while. Transferring data and doing some basic calculations will help him spot omissions that have slipped through. Since the tabulations are done shortly after the data have been gathered, it is often still possible to return to the farmer and obtain the missing information or to correct errors.
In addition to these simple checks for completeness and plausibility, the forms can point out possibly incorrect data.

For example: In some surveys, the areas of fields are not measured but are simply estimated by dividing their production by the yield per ha as obtained on sample plots. When these areas are entered on the recapitulative forms, it may appear that one farmer has sown an exceptionally large area with maize: 2.0 ha as against the range of 0.1 to 1.3 ha of the other farmers. This would seem to imply that this farmer has a much larger farm than the others, but if a check is made against his labour force and this is found not to be exceptionally large, then something is amiss.
The supervisor should first check the obvious causes of error:

whether the enumerator wrote 0.2 as 2.0, whether he wrote the area for white sorghum instead of maize, or whether he made a mistake when calculating the area. Perhaps the field is a communal field worked by a youth group and the enumerator has filed it under the name of their leader. If none of these applies, the dimensions of the field should be checked. Pacing the field does not take much time and is sufficient to show whether the area estimate is roughly correct. If it turns out that the field is indeed very large, this should be noted on the form, together with any reason the farmer has for sowing so much maize.

7.2.1 Layout of recapitulative forms

Recapitulative forms should be kept simple and easy to fill in in the field. Data should be entered on them in such a way that the data can be used directly for the analysis (or keypunching) without further conversion.

There are some general rules for the layout of recapitulative forms that will make them easier to use at the office later. These are:

- The farmers should always be listed in the same order, with the same amount of space given to each;
- To distinguish groups of farmers or types of farms, paper of different colours can be used;
- If subgroups are to be selected within the sample, it might be advisable to put the recapitulative tables on 'edge-notch cards'. (These will be described in Chapter 8.)

7.3 Feedback from field to office

At regular intervals, if possible every two or three weeks, all the supervisors should come to the office for a meeting with the director. They should bring with them any completed survey forms and also the recapitulative forms they are updating at that time. The unit director can go over the forms to check the work of the supervisors and enumerators and to keep himself informed of the changing situation in the field.

At the meeting, the supervisors should discuss their work, reporting any unusual situation they may have encountered and how they handled it. If they have met with situations they could not handle, this too should be discussed and a decision made on how such situations should be dealt with in future. In this way, the supervisors can make sure that they all give the same instructions to their enumerators.

If a supervisor brings some survey or recapitulative forms containing data which he has questioned but could not verify in the field, he should discuss this with the director. Together, they can decide what should be done about those forms.

During their meetings, the director and supervisors can discuss the program for the next few weeks, deciding on the schedules for new surveys or for training sessions, if required.

7.4 Selection of data for analysis

Some standards should be set to determine which data will be accepted for analysis and which will not. Ideally, only complete sets of data should be analyzed. Often, however, so many essential data are missing that the standards have to be less demanding than the ideal. In whatever way it is decided, a limit must be set on what is usable and what is not, so that at least the data from different farmers meet the same standard.

For example: A labour survey done by repeated interviews is not complete, because a few interviews were not done for one farmer during the weeding period. This data set cannot be used to estimate that farmer's total labour input per crop. But if the rest of the period is well documented, the data on other activities (e.g. sowing and ridging) can still be used.

Data for farmers in different subgroups should be at a similar level of accuracy if the data are to be analyzed together.

For example: The yields of cowpeas are being surveyed for several groups of twenty-five farmers each, in different climatic zones. The enumerator of one group was ill during the cowpeas harvest so he missed most of the harvests and got only four reliable estimates. If the enumerators in the other locations got the yields of their twenty-five farmers correctly, the four estimates of the first enumerator will not be considered for the analysis as they will lower the quality of the overall result. If, however, the other enumerators also had problems and got only five or six yield estimates, the four estimates of the first enumerator can be used, although the analysis then has to be kept simple because the data base is so small.

After such checks for completeness, the data sets available for analysis will have been reduced. Sometimes an otherwise good set of data will have been eliminated from a specific type of analysis. This is hard to accept. But it should always be remembered that the quality of the results can never be better than the quality of the data used, and that this is limited by the quality of the worst data. Cleaning up the data before beginning an analysis can only improve the value of the final results. When these results are being reported, the number of data sets actually used in the analysis should be mentioned so that the figures cannot be misinterpreted by the users.

The data sets remaining after the elimination processes will be re-grouped and analyzed, but this does not mean that the checks on data are finished. During the entire process of tabulation and analysis the data will be checked continuously. It is the responsibility of the data processors to notice any strange numbers or odd situations and to put these forms aside to be checked again, if possible in the field or with the supervisor. This can lead to revisions of what data are usable.

8 DATA PROCESSING FOR MANUAL ANALYSIS

In many monitoring and evaluation systems, data processing is a badly neglected activity. Most of the effort is put into the collection of reliable data - as many as possible within the limitations of personnel and material. This disequilibrium causes long delays in data processing and ultimately reduces the analyses that can be made.
Independent of the calculation tools available (calculator, micro-computer, or regular computer), some of the data will always be processed by hand. This chapter will discuss how to process and analyze data by hand, placing emphasis on methods that get the basic results of the survey to the users as quickly as possible. Rapid reporting can bring swift reactions from the users, so that their programs can be adapted within one crop season. In semi-arid regions, with only one crop season a year, a delay in data presentation often means that an entire year is wasted before changes can be introduced. With on-going tabulations on recapitulative forms starting right from the beginning of the farm surveys (Chapter 7), interesting developments can be reported almost as soon as they are noticed.

8.1 Tabulations for rapid reporting

The layout of a survey form is always a compromise between different objectives:

- The form should be easy for the enumerator to use, both when asking the questions and when writing down the farmers' answers;

- The wording and sequence of the questions should be clear and logical to the farmer so that he understands what is being asked;
- The form should make it easy to tabulate the results and to transfer the data to recapitulative forms. This, however, must not conflict with the above two objectives.

Processing the data

Recapitulative forms have the same layout, whether the data are processed manually or by computer. The recapitulative forms discussed in Chapter 7, as a means of checking the quality and completeness of the data in the field, can also be used at the office for the first calculations of averages, dispersions, and frequencies of observations. If some of the data are still missing, the results can be presented as preliminary. The supervisors should keep the recapitulative forms until they are complete, but should bring the forms with them when they visit the office for their meetings with the director. Data from the forms can be photocopied, if need be, for further use in the office, unless a notebook with carbon copies was used for the forms in the first place. At the office, the forms can be completed by calculating the averages and totals in the extra columns reserved for that purpose. In this way, the data are soon available for presentation in the regular reports of the unit.

For example: If the extension service has advised the use of fertilizer on a certain crop, they can be informed of the success or otherwise of their program almost as soon as the period for fertilizer application has ended. With such rapid reporting on early crop activities, the extension service can adapt its strategy for later crop activities.

The table presented in Chapter 7 (Table 7.1) can be extended to include the use of animal traction, thinning of the crop, or any other topic of interest to the extension service (or to any other user). Those data, together with estimates of the data not yet available, provide useful feedback about the success of recommendations.
If the amount of data gathered on a crop is not large, it can be transferred to one or two recapitulative forms, which will be easy to verify and can later be used directly for much of the analysis, and even directly coded and keypunched if a computer is used. When Table 7.1 has been completed, for instance, it is a simple matter to add up the yields or production figures for all the farms and to calculate the average and the dispersion for that group of farmers.
For data gathered during a single-visit survey, or on forms sent back to the office each month, the tabulation and basic calculations can both be done at the office.

For example: A survey has been held to make an inventory of the agricultural equipment on each farm in a sample. The survey covered the type, age, and purchase price of each piece of equipment. The entire survey was completed in two weeks and, after the forms had been checked by the supervisor, they were sent to the office to be analyzed. After the data had been checked again, the results were calculated and entered on a recapitulative form. For this survey, only the results of the calculations were entered on the recapitulative form, because the original inquiry forms were filed and are thus permanently available should further checking be desired later

Table 8.1. Example of a recapitulative form

Name of farmer .... Location ....

GENERAL INFORMATION (A)	Years			
	19..	19..	19..	19..
Number of workers				
Number of labour units				
Number of consumers				
Total area cultivated (ha)				
Area per labour unit (ha)				
Net profit per worker (CFA)				
Net profit per unit of labour (CFA)				
Net profit per consumer (CFA)				
Net profit per ha (CFA)				

TOTAL FARM INCOME (B)	19..	19..	19..	19..
Gross production value (CFA) from F				
Indirect costs (CFA)				
Net profit (CFA)				
Income from other sources (CFA)				
Total farm income (CFA)				

INVENTORY HERD (C)				
Type	Number at counting			
	19..	19..	19..	19..
Cattle Male 0-<2 years				
over 2 years				
Female 0-<2 years				
over 2 years				
Donkey				
Horse				
Sheep				
Goat				
Pig				

INVENTORY AGR. EQUIPMENT (D)				
Type	Number at counting			
	19..	19..	19..	19..
Plough				
Triangle				
Harrow				
Long yoke				
Short yoke				
Plough share				
Weeder share				
Ridger				
Cart				

NET RESULT BY SUBJECT (E)				
Subject	Net results (CFA)			
	19..	19..	19..	19..
Agr. equipment				
Small agr. tools				
Animal traction				
Rest herd				

Table 8.1. (contd.)

AGRICULTURAL INFORMATION (F)	Crops in year 19..							Crops in year 19..						
Yield (kg/ha)														
Area cultivated (ha)														
Value of crop production (CFA)														
Direct costs: seeds (CFA)														
fertilizer (CFA)														
insecticide (CFA)														
Gross margin (CFA)														
Gross margin per ha (CFA/ha)														
	Crops in year 19..							Crops in year 19..						
Yield (kg/ha)														
Area cultivated (ha)														
Value of crop production (CFA)														
Direct costs: seeds (CFA)														
fertilizer (CFA)														
insecticide (CFA)														
Gross margin (CFA)														
Gross margin per ha (CFA/ha)														

on. The results were transferred to a table and promptly reported to the users.

8.2 Tabulations for further analyses

Survey results presented in tables serve as the permanent data base for analysis. These tables regroup all the data recorded on the original survey forms and on the recapitulative forms completed either by the supervisors or by the office staff. The original forms are then filed away and no longer used except if, in case of doubt, data need to be checked.
This section will show how tabulations can aggregate data at any level likely to be needed for analysis. As an example, we shall present a form on which all the data gathered about one farmer over a period of four years can be tabulated.

Example of a recapitulative form for one farmer

Table 8.1 represents the front and back of the recapitulative form. The name of the farmer and the location of his farm are entered in the heading of the form. Part A provides space in which to enter information about the members of the household. The number of workers means the number of people who have worked to produce crops in that year. This can be adjusted for the sex and age class of each worker to obtain the number of labour units. The labour units take into account the fact that a woman is usually not available full time for agricultural work (she has to prepare meals, fetch water, etc.) and that a child cannot do as much work as an adult. The number of consumers and the total area cultivated are needed to calculate the ratios included in the table. (These ratios provide an indication of the viability of the farm; ratios and the other economic terms used on the form will be discussed in Chapter 10.)
Part F on the back of the form regroups all basic data per crop (the aggregation of different fields) in monetary terms. The gross margin (the value of production minus direct costs) for each crop can be

calculated directly on the form. Aggregating the gross margins of all crops provides the gross production value of the farm which is entered in Part B. The net profit of the farm is then calculated by subtracting the indirect costs. If the off-farm income is known, the total farm income can be calculated in Part B. The form includes space for revising the prices used in the calculations, as these can fluctuate and they must be kept comparable over the years.
Part C presents an inventory of the farmer's livestock, and Part D an inventory of his agricultural equipment. Both serve as background for the interpretation of Part E, in which the income generated from the herd is calculated and also the costs incurred for agricultural tools and animal traction, which are the basic elements in calculating the indirect costs.
The extra space makes it possible to mention costs (or benefits) not included on the form.

This form is merely an example of what can be done. Many other such tabulations can be made. They can present figures for one year only, if the data base per farmer is large, or can tabulate figures per individual field.
The tabulations should be stored and kept accessible at all times, because they will be used often to check the data, perform analyses, or be updated. If at all possible, the tabulations should be done on cards rather than on paper.

8.3 Sorting strips

The agricultural activities of an entire farm are monitored so that the interactions between the different activities can later be analyzed. Some interactions are expected but others are not. One therefore has to examine different combinations of variables together. Often the number of variables that have to be combined in various ways is so large that they cannot all be put into one table. Yet, it is very difficult to discover interactions between variables if they are not placed side by side. A solution to this problem is to use sorting strips. These allow different combination of variables to be examined without having to copy

the data time and again.
A sorting strip is a piece of firm paper on which data on one variable only are recorded. It corresponds to one column of a recapitulative form. If the information for each farmer (or field) is always put on the same line on the strip, the strips can be placed side by side in different ways to form different tables. Sorting strips are only useful if no further stratification or other grouping of data is to be made. They can be practical and time-saving tools for analyses by hand. Care should be taken that each strip clearly indicates which variable is put on it, and that all strips are carefully filed; small pieces of paper are likely to get lost.
The strips can easily be made by drawing lines on a sheet of light cardboard and cutting it into equal columns. Colour coding for different stratifications in a sample can be of great help.

8.4 Edge-notch cards

Edge-notch cards are practical for the rapid identification of those elements in a sample (farms, fields, or whatever) that have one or more characteristics in common, each element being represented by its own card.
Edge-notch cards are thin cards with a corner cut off to make sure that the cards are all stacked in the same way. On each card, one or two rows of holes are already punched along the side (Figure 8.1). Each pair of holes is identified by a printed number or letter and each hole can be made to represent one characteristic.

For example: In Figure 8.1 the outer hole of position 12 can be said to represent the variable 'farmer's ownership of a plough'. If a farmer does not own a plough, that hole on his card is left untouched; if he does, the hole is clipped open, as is shown in the figure. When the cards have been prepared for all the farmers in the sample, it becomes easy to separate the cards of the farmers with ploughs from those without. The cards are stacked, all with the cut-off corner together, and a knitting needle or bicycle spoke is put through the outer hole number 12. When the pack is hung from the needle, the cards with an open hole 12 (the cards of the farmers who own a plough) will fall off. The cards where hole 12 has been left untouched (the cards of the farmers who do not own a plough) will remain on the needle.

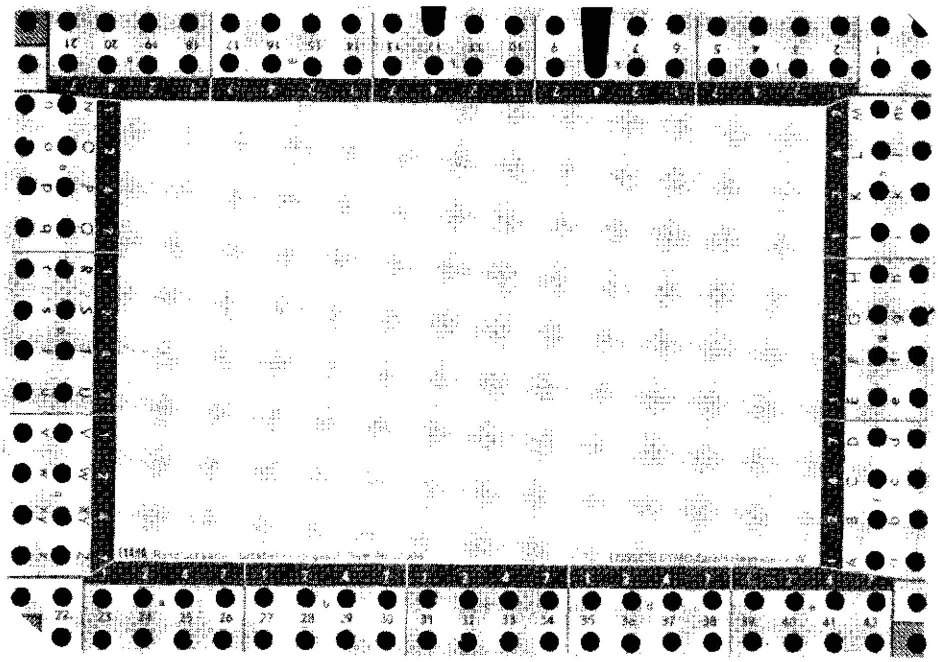
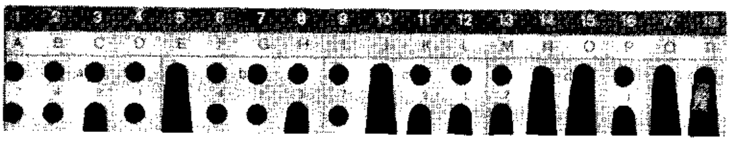


Figure 8.1. Example of an edge-notch card

The same system can be used for variables with more than just a 'Yes' or 'No' alternative. Examples of how to indicate a variable with more alternatives are presented in Figure 8.2.

Variable with three possible alternatives, using one pair of holes



Variable with four to nine alternatives, using two pairs of holes

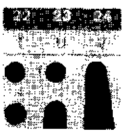


Figure 8.2. Coding of edge-notch cards

Convenient as edge-notch cards are, they have some disadvantages:

- They are not always locally available (but can possibly be ordered through an office supplies store);
- They have to be used carefully because the holes are easily damaged (especially with frequent use);
- As a tool for analysis, edge-notch cards are limited by the number of alternatives that can be punched.

Variables with many different values can still be considered if the values are grouped into classes. The holes of the cards cannot be used to indicate the farmers' agricultural income as such, for there are too many different incomes; but if the farmers' incomes are grouped into, say, five classes, two pairs of holes are enough to represent the variable. If the actual values of the data are then written on the cards, the cards can be used as recapitulative forms as well.

Example of an edge-notch card as recapitulative form

Several farm surveys have been conducted to gather information on the main aspects of crop cultivation for a sample of farmers. The data are to be analyzed by hand. To be able to group the farmers according to different criteria, the data have been entered on edge-notch cards, one card per farmer. The surveys started in 1975 and were to be continued for three years.

Figure 8.3 represents the two sides of an edge-notch card containing the basic information about a farm for one crop season (Part A), the basic production data (input/output) for each field (Part B), the gross margin per crop in monetary values (Part C), and the gross production value per labour unit and per hectare (Part D). The most essential information - that by which subgroups are to be selected - is coded by hole-punching at the edge of the card. Here the position 73 (first pair of holes at top left) represents the crop season described on the card. The survey lasted three years so that one pair of holes is sufficient. (Code applied: 1975, holes untouched; 1976, one hole clipped; 1977, two holes clipped.) Next, the geographical location of the farm is coded. Four villages were surveyed, so two pairs of holes (72 and 71) are necessary. (Code applied: Village I,

Year 1977 Name ..Peter...S:....ID.74.29.070...
Farm type...12.......Equipped 1975 Labor Units.3.75..
Area Cultivated, Total..7.4.ha, per labor unit.1.9..ha

	Cotton I		Cotton II		Maize					
	/ha	T	/ha	T	/ha	T	/ha	T	/ha	T
Area (ha)	—	1.2	—	0.7	—	1.9				
Product.(kg)	1560	1,872	1,200	840	1,430	2,717				
Seeds (kg)	20	24	20	21	20	38				
Fertil.(kg)	130	156	70	49	—	—				
Insect.(l)	9	10.8	13	9.1	—	—				
Labor (CFA)		—	100	70	—	—				
Gross Margin	[illegible]	[illegible]	[illegible]	37,549	[illegible]	136,[illegible]				
(CFA)										
Techniques										

	White Sorghum		White Sorghum 5		Red sorghum						
	/ha	T	/ha	T	/ha	T	/ha	T	/ha	T	
Area (ha)	—	1.2	—	1.0	—	1.1					
Product (kg)	670	[illegible]	770	740	[illegible]	[illegible]					
Seeds (kg)	5	6	8	8	10	11					
Fertil (kg)	—	—	—	—	—	—					
Insect (l)	—	—	—	[illegible]	[illegible]	—					
Labor (CFA)	—	—	—	[illegible]	[illegible]	[illegible]					
Gross [illegible]	36,575	43,[illegible]	40,260	40,260	[illegible]	55,252					
(CFA)											
Techniques											

Gross Production Value (CFA) [illegible]

GPV / labour [illegible] 107,845

GPV / ha 56,[illegible]

holes untouched; Village II, first hole clipped; Village III, first pair of holes clipped; Village IV, second pair of holes clipped.) In the same way the experience of the farmer with animal traction can be indicated using the variable 'year equipped' as indicator. Not more than six different 'years equipped' were observed so this variable, too, can be clipped by using two pairs of holes: 70 and 69. (Code applied: not equipped, holes untouched; equipped in 1973, one hole of 70 clipped; equipped in 1974, both holes of 70 clipped; equipped in 1975, first hole of 69 clipped; equipped in 1976, both holes of 69 clipped; equipped in 1977, first hole of 70 and 69 clipped.) The size of the labour force, a variable with many possible alternatives, can be clipped only after the alternatives are grouped into categories. A classification into three categories was considered to be sufficiently precise and therefore one pair of holes (68) is needed on the card. (Code applied: 0 - < 2, holes untouched; 2 - < 4, first hole clipped; 4 and more, both holes clipped.) In the same way, many other important variables can be clipped (e.g. income, use of fertilizer, etc.).

A card like this can be prepared for each farmer in the sample. By stacking the cards and putting a knitting needle through a hole, it is easy to separate the subgroups and compare their results.

For example: The cards of farms with a large labour force can be extracted to find out whether they are worked more intensively and whether they have a larger gross income per unit of labour.

8.5 Some considerations for analysis by hand

Using a calculator and performing an analysis by hand has important advantages over computer analysis:

- The entire data processing and analysis can be done by the staff of the unit. They are aware of the situation in the field, so if some odd information is found, which is quite likely, even after the many checks already done, they are able to recognize it and verify its validity. This kind of continuous checking throughout the analytical process depends on human knowledge and experience and cannot be replaced by machines;

- The type of analysis actually done is limited by the quality of the data and, here, also by the knowledge and experience of the unit director. He is only capable of doing an analysis which he completely understands. This reduces the risk of misinterpretation of the results of the analysis;
- A first simple analysis by hand can be finished earlier than an analysis done with a computer. No matter how the work is done, the data have to be checked and transfered to recapitulative forms. Doing some basic analyses (totals, means and dispersions) from the raw data on recapitulative forms gets much faster results than keypunching the data, checking the keypunching, and then awaiting the results of analysis from the computer;
- Analysis by hand can be a rather lengthy process. This is not necessarily bad if it encourages a thorough consideration of whether the analysis is worth doing before the work is started. This avoids the presentation of results of an analysis that does not mean anything but just happens to correlate by chance (so-called nonsense correlation, something that can easily happen in computer work).

A disadvantage of analysis by hand is that more data processors may be needed if the data base is large. The smaller the staff assigned to data processing and analyses, the longer it will take before the results are available, which reduces the usefulness of the results. But, as mentioned before, this can be avoided by properly allocating the resources of the unit over data gathering and processing right from the start.

9 DATA PROCESSING FOR COMPUTER ANALYSIS

In theory, the type of analysis that can be performed on a body of data depends only upon the validity of the data, not on the tool used for the calculations. In reality, however, some calculations are simply too cumbersome to be done, especially when one is working with large samples. With a computer, a very large body of data can be handled and complex calculations can be performed quickly. This is true even of a micro-computer.

Nevertheless, even with a computer, the main limitation in analyses always remains the validity of the data, which still have to be checked and tabulated as described earlier. The data then have to be keypunched and checked again before being fed into the computer. This process, and the advantages and disadvantages of manual versus computer analysis, will be discussed below. First, however, a few of the key terms used in computer work will be explained.

9.1 Definitions

Program

A calculator does not think, it merely follows instructions. A set of instructions describing - step by step and in standard terms - what a calculator has to do to make the required calculations is called a program. A simple calculator has the basic programs (addition, multiplication, square root) built in as part of its standard equipment. Programmable calculators have many programs already built in and can do more complex calculations if they are given step-by-step instructions.

Computers, which are actually very large calculators, only operate if given a program. Writing a computer program is no easy task; it requires knowledge of the computer and of the language that is to be used so that the computer 'understands' the instructions. It takes some experience before one is able to use even the simplest language, and most computers require languages that are quite complex. The analysis of data from farm surveys often needs complicated programs, which even an experienced programmer will find difficult to write. For that reason, computer companies have prepared standard programs, which an experienced programmer can adapt to the specific needs of a user. Because a program is so complex, it is very unusual if it works during its first run. It has to be tried, revised, and tried again several times before it is ready for use.

Coding

Data have to be coded before they can be fed into the computer. Coding can be numerical, with each item of information represented by a digit, or it can be alphabetical, with each item represented by a letter. Either system, or a combination of the two, can be used in computers.

Keypunching

To run a program, the computer needs to be given the data that are to be analyzed. This is done by copying the coded data from the forms or tables using a special sort of typewriter. This is called keypunching the data. The data are thus transferred to punch cards, magnetic tapes (as in tape recorders), or disks, which the computer can read.

At the same time the data can be printed on paper (called a print-out) so that the keypuncher can check what he has entered.

Memory

A computer's memory is its ability to store information after that information has been entered, so that it can be used at a later phase of the calculation without having to be copied and entered again. Calculators have a memory too, but it is small and can store only a few figures at a time - and usually does this only as long as the machine is not turned off; when it is turned off, everything is erased. Computers can retain a large quantity of information in their memories, where it remains available throughout the calculation. It can also be stored for later use (see below).

Storage

The memory of a computer only retains the data as long as is needed to complete the work described in a program. Basic data or results that will be used more than once therefore have to be stored somewhere so that they can be recalled, i.e. put back into the memory, ready to be used again without the need to be copied again. This is possible with computers, as they can store the information on punch cards, magnetic tapes, or on disks.

File

For most of its calculations, the computer uses only small portions of the data at a time.

For example: To calculate a farm's gross margin from cereals, the computer adds up the gross margin of all the fields with a cereal crop (according to a pre-set list), but not the gross margin of other fields. To calculate the gross production value of the farm, however, it will add up the gross margin of all the fields.

In general, the data are keypunched into the computer in groups of variables which will be needed together for analysis. The data entered as one group of variables is called a data file. It is possible to do an analysis by combining data from different files, but it is not always possible for the programmer to combine different files satisfactorily.

Some standard programs can only use one type of file structure. So before the data are keypunched, one should decide which analyses will be done and then select the file structure compatible with these analyses. In this way unforeseen file combinations to be done later are kept to a minimum.

9.2 Computer analysis

Theoretically, a computer is an ideal medium for analysis because it has a large memory capacity, can perform any kind of calculation, and can store the data for later use. It is extremely fast in performing even complex calculations and can handle a large data file. On the other hand, it has certain disadvantages, which should not be underestimated. These will be discussed below.

9.2.1 Preparing for analysis

It is a commonly-held belief that running an analysis on a computer will solve all the problems of analysis; once the data have been fed into the computer, it is merely a matter of pressing the right button. But this is just not true! While calculations can indeed be done very rapidly, there is more than that involved; preparing for calculation can be extremely time-consuming.
The data to be entered from the forms or tables first have to be coded and then keypunched. The keypunched data then have to be checked and errors corrected. This can be a lengthy operation because the keypunchers are usually not aware of the meaning of the figures they keypunch and can therefore make the strangest errors.
Sometimes the programs by which the data are going to be analyzed have to be written because standard programs do not provide the desired analysis; in the best of circumstances, standard programs will at least

have to be adapted. The unit director has to explain to the programmer what he really wants. Director and programmer, both specialists, talk from entirely different points of view so it is not easy for them to achieve complete understanding. When the unit director is not familiar with computer work, even defining his needs to the programmer can be difficult. Especially at first, it can take a long time to develop a program.

The data have to be entered on file structures suitable for the analysis, a process that should be followed right from the beginning. Otherwise very complex programming will be required just to retrieve the data needed for analysis, which then have to be filed differently from the way they were first filed.

The steps outlined above represent weeks, or even months, of work.

9.2.2 Costs

A computer is very expensive and few projects could afford to buy one or to employ the staff needed to use it. Most countries these days, however, have specialized computer centres which will rent their services to a project for a fee. The costs are likely to be high - particularly if special programs have to be written - but these costs are more likely to be justified if the same programs are to be run routinely on a large body of data over several years.

When asking for a price estimate from a computer centre, one should be able to tell them as precisely as possible the quantity of data involved; the computer staff can then estimate the amount of work required. A simple way to approximate the quantity of data is to count how many different items of information will be collected for each element in a sample, including all the different data needed to define and describe each element. This number is then multiplied by the number of elements in the sample. Even a simple survey is likely to produce many thousands of items of information.

For example: The recapitulative form per farm presented in Table 8.1 may contain 400 different items of information (rows × columns + title). On a sample of 200 farmers, that amounts to 80,000 different items, and this is just one form!

The disadvantages of using a computer centre, in addition to the costs, are that it may entail delays because its staff works for many clients. Technical breakdowns may also cause delays because not all countries have repair staff and parts readily available.

9.2.3 Personnel constraints

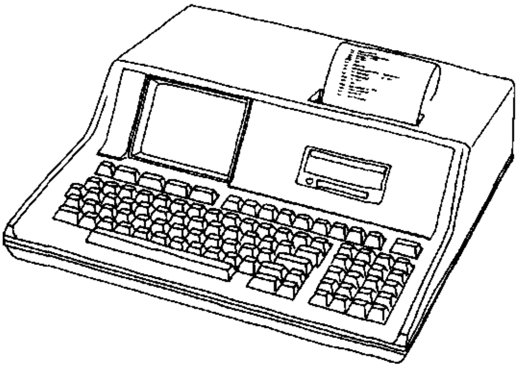
Computer analysis is completely out of the hands of the monitoring unit, so no direct personnel constraints are involved. But the unit has to rely on a computer programmer to prepare the programs and decide how the data should be keypunched. If the programmer is not familiar with farm surveys, it may take many meetings between unit staff and computer experts and many revisions and adjustments to the programs before they are ready for use.
There is a risk inherent in all computer work: because of the ease and speed of calculations, it is always tempting to run complex analyses which are technically correct but are not appropriate to the quality of the data base. If a professional programmer is involved, the unit director may find it difficult to explain to him what kinds of analyses ought to be done - in the light of the quality of the data - and what data ought to be used.

9.3 Analysis with a micro-computer

A micro-computer is a small table-model computer whose memory is smaller than that of an ordinary computer but which costs much less. A micro-computer comprises:

- A keyboard (like that of a typewriter), which is used to enter the information into the machine;
- A screen (like that of a TV set) to display the information entered and the results calculated by the machine;
- A printer to print out, on paper, the information entered and the results obtained;

- A system to store data or programs so that they can be recalled and used again. Depending on the machine, the data and programs are stored on disks or on magnetic tapes (cartridges or cassettes).



A micro-computer

A micro-computer works like a regular computer but it has a smaller memory and is therefore of limited use for analysis. Most micro-computers use programs written in Basic Language, a computer language easier to learn than the regular computer languages. Pre-recorded programs are available for the most usual calculations. These machines seem to offer interesting possibilities for development projects but they are fairly new and few projects have yet used them.

9.3.1 Material requirements

A micro-computer is sensitive to high temperatures, dust, and a dry atmosphere, so it usually has to be kept in an air-conditioned room and covered with plastic when not being used. It requires a continuous supply of electricity. If the supply is interrupted, the memory of the machine will be automatically erased, and the operator has to start all over again. Some types have a constant memory, the information remaining in the memory even when the power is turned off. The voltage should be constant, so in many countries it is advisable to use a voltage regulator to cut off the peaks in power supply which would otherwise damage the machine.
Supplies, spare parts, and maintenance may be difficult to obtain. A

large stock of supplies (printer paper, storage cassettes, cartridges or disks) should be kept at the office. It may be wise to keep an extra machine as a spare, should one break down. Some are cheap enough for that.

9.3.2 Personnel requirements

To use a micro-computer to its full potential requires some programming abilities, which are best acquired by following a training course. If the operator of the micro-computer has no programming experience, he can, with some practice, use the machine as a large calculator, applying only the standard programs.

9.3.3 Advantages and disadvantages

The micro-computer has one great advantage over calculators, even when used by people who are not capable of writing programs for the machine. With a standard program, it can analyze a large body of data and can store the data for later use. Standard programs for micro-computers contain more types of analyses than do standard programs for calculators. There is a risk, however, that the operator performs analyses that he does not quite understand, and may therefore misinterpret the results.

Used by people who are able to program the micro-computer themselves, this tool opens many prospects. With very little coding, the data can be stored on permanent files from which specific data sets can be selected for analysis with a program well adapted to the situation. Here, there is little risk that the operator will perform an analysis he does not understand, since he must first write a program for it. A disadvantage of this system is that if the programs have to be designed and developed during the process of analysis, it may take a long time before the programs will run smoothly. During this design and development stage, data processing by hand has to be continued.

The micro-computer is a promising tool for the analysis of farm data, especially as more ready-made programs become available. It is relatively easy to use as a (large) calculator on which large bodies of data can be entered and analyzed. But to use it as a small computer on which the data are entered and stored, to be recalled as desired for later analyses with special programs, is not possible without a thorough knowledge of the programming abilities of the machine.

9.4 General recommendations

In most projects, the use of a regular computer in monitoring and evaluation of farm data is not advisable, especially in the beginning. The more work that is done by hand in the beginning, the more chance that errors will be detected. When a smooth routine for analysis by calculator (or micro-computer) has been established and the data base is gradually becoming so large that it cannot be handled that way for much longer, computer analysis can be started, but hand analysis should be pursued at least until computer analysis runs smoothly. It may be best to keep on performing basic calculations by hand for rapid preliminary reporting.

If the volume of data is expected to reach such proportions that computer work seems necessary, or if the unit intends to buy a micro-computer, it is wise to request technical assistance in the form of a consultant computer programmer. The unit will not need the services of a programmer permanently, so an efficient, less costly arrangement could be for a consultant to come only when needed. Ideally, the consultant should be available for several weeks when the work program of the unit is being prepared. He can then participate in the design of the data-gathering forms and the recapitulative forms. He can check that the plans for processing the data are adapted to easy keypunching, with as few transfers of data as possible. Whatever coding may be required should be planned at that time. The consultant can prepare tentative programs for the desired analyses, or adapt standard programs as needed. He can also conduct training sessions for the unit director and tne data processors.

Later on, when there are enough data checked and keypunched to try the programs, the consultant should come back to review the programs, check the work of the unit's staff, and provide any additional training that

might be required.
Even with properly set up computer programming, however, the unit director would be wise to continue to have the on-going tabulations done by hand to ensure better validity control and to make rapid preliminary reporting possible.

10 EVALUATION OF HOUSEHOLD INCOME

As was explained in Chapter 3, the program of data gathering and analysis is more than just processing the data as they happen to come in. Data are gathered and analyzed as a function of the type of evaluation desired. Whether that evaluation focuses on a situation at one point in time or on a process of agricultural change, it will always involve judgements on agricultural policy and recommendations on how that policy might be adapted.

Agricultural change can be evaluated either at farm level or at project level. An evaluation at farm level focuses on the development of the farm throughout the years within all its constraints. An evaluation at project level focuses not on the development of individual farms but on the development of the project as a whole, in the context of the regional or national economy. The two types of evaluation can have contradictory results: a positive development at farm level can co-exist with a negative development at project level, and vice versa. The work of the monitoring unit will usually be limited to farm evaluations. The project itself will be evaluated by outsiders, using the unit's data.

A farm evaluation takes the farm as the central unit and analyzes how certain aspects of the farm are changing or how the farm as a whole is developing, especially in economic terms. Key elements of such an analysis are:

- Results from crop cultivation;
- Farming practices;
- Income from livestock;
- Off-farm income;

- Total household income.

The evaluation always first considers these elements separately and then, if possible, together, because a change in one element may influence other elements.

For example: The introduction of new varieties has improved the yields of cereals but, because the new varieties require a greater labour input, the off-farm income has decreased, leading to a lower total income. Such an 'improvement' is not desirable for the farm household.

10.1 Context of the evaluation

10.1.1 The socio-economic environment

The results achieved at farm level can only be interpreted correctly if placed against the socio-economic environment in which those results were achieved.

For example: If a farmer harvests 1,000 kg of maize per ha, it is difficult to evaluate whether this is much or little - and therefore whether the situation could be improved - unless one also knows the environment in which that yield was achieved.

For this reason the basic characteristics of agriculture in the region have to be reviewed and the results of the evaluation placed in that context. This gives the user a better understanding of the information obtained.

The characteristics that need to be reviewed are the following:

- Climatological conditions of the area (temperature, rainy season, rainfall, evaporation, etc.);
- The local population (number of inhabitants, number of workers per household, etc.);
- Traditional agricultural practices (cultivation techniques, inputs used, yields achieved, etc.);
- Improved practices being introduced.

10.1.2 The farm

In the semi-arid zones of West Africa, the most common farming system is the subsistence farm, so most agricultural improvements are aimed at that type of farm. To the outsider, a subsistence farm looks like one entity. It is cultivated by the (extended) family, and is represented to the outside world by the head of the household, who is almost always a man. He is the person who goes to the agricultural extension meetings, who has access to credit, etc. But the farming household is usually far more complex than this. The head of the household controls only part of the farm, the communal fields. These fields are worked by all members of the household, with the head of the household controlling operations. The other fields of the farm are personal fields, fields that belong to individuals in the household; most adults have their own personal fields. These fields often receive a higher labour input than communal fields do, because the person providing the labour is more directly interested in the results. And yet, the individuals do not always have access to inputs such as fertilizer or high-yielding varieties of seed.

The complexity of ownership is often ignored in project design, improvements being introduced only to the head of the household. In an analysis comparing the yields of a crop with different input levels, the elements of the population must be homogeneous. If the personal fields are very different from the communal fields and both are lumped together in the analysis, few meaningful results can be obtained. This should be kept in mind, especially when the following subjects are being analyzed:

- Assimilation of agricultural extension recommendations;
- Relation between inputs and yield;
- Crop rotation;
- Management and decision making.

10.1.3 Technical package being promoted

If agricultural development is being promoted by the introduction of new farming techniques, the evaluation will include a comparison of the farmers' traditional practices with the techniques being promoted. This means that one must have a detailed description of the new techniques, something that is not always easy to obtain. In many projects, while the project is being implemented, the extension service modifies its original program. These modifications are disseminated only through memos; no updated description of the entire package exists. It can then be difficult to find out exactly what the extension service is really promoting. And yet, without a clear statement of the standards against which the activities of the farmers are being evaluated, the evaluation can have little meaning.

Extension agent introducing a new crop

10.2 Assimilation of agricultural innovations

The first element to be analyzed in an evaluation of farm results is whether the farmers are assimilating the proposed agricultural innovations. Are they following the advice given by the extension agents?

In reality, it is rare to find that the farmers accept or reject an innovation in its entirety. One must therefore look at the degree to which the innovations are being assimilated. This means that the entire technical package has to be divided into small units for which the frequency of assimilation can be calculated.

For example: If the extension program is promoting the use of 200 kg of fertilizer per hectare for sorghum, it should be examined how many people applied no fertilizer, how many applied 0-<50 kg, how many applied 50-<100 kg, and so on.

By examining all the key elements in the package in this way and then combining the results, it is sometimes possible to identify the farmers with a relatively high or relatively low degree of acceptance of the new technology. The results of these two extreme groups can be analyzed and compared. This might lead to a better insight into some of the constraints that impede farmers from accepting the new technology, choosing as they have to do from among conflicting options.

Some of the factors that might constrain the farmer are the following:

- Availability of money: innovations usually cost money, a rare commodity for a subsistence farmer. He may not be able to afford the innovation or, if he can, may decide that investing in an expensive new technique is just too risky;
- Availability of time: most improvements in crop cultivation mean the allocation of more time to the crop. Technically, the family might have that time available but the sacrifice of leisure time, or in paying social visits, or time spent on other activities might be considered too big in relation to the expected improvement in crop production, 'When', additional work is required is as important as 'How much';
- Availability of inputs: transportation or other problems might cause a shortage of necessary inputs;
- Know-how: the farmer might have problems in understanding the new techniques. The professional knowledge of the extension agent and the frequency of his visits can help the farmer;
- Who benefits from the improvement? Some improvements may be beneficial only to farmers who have access to irrigation or who own

expensive equipment;

- Other considerations such as: Will the surplus be easy to sell? (The taste of a new variety can be decisive.) Does the cropping calendar conflict with tradition?

These are only a few of the more obvious constraints that might influence the farmer's choice. Often, constraints are not easy to identify and have to be researched in depth.

Analyzing the weak points of the extension program provides information not only about what was not well assimilated by the farmer but also why not. It is not an easy task to find out why a farmer does not follow the advice of the extension agent. It is tempting to simplify by looking only at economic factors, which can indeed be major limitations, but are not the only ones. Factors such as who has the right to sell the crop, the area that has to be cultivated with the staple crop to ensure self-sufficiency in food, and the taste of a new food crop can be just as decisive for the farmer in his selection of farming practices.

Analyzing the extension program in this way will also identify the strong points of the program, showing where the extension effort might be reduced in the future.

The introduction of new farming techniques is likely to have been based on economic considerations, taking the other factors into account as constraints if they are known. An economic analysis of the results achieved in reality in comparison with what had been expected can therefore show whether, from an economic point of view, the farmer was right in his decisions.

10.3 Crop results

10.3.1 Yields

Data about inputs (means of production required to produce a crop) and outputs (crop production) need not always be expressed on the basis of a unit of land. When the influence of early sowing or that of fertilizer use is to be analyzed, however, data must be expressed on a 'per hectare' basis. The same applies when one wants to compare the production of fields of different sizes, which must be expressed as yields (kg/ha).

The analysis of the yields is begun by grouping all yields of a certain crop into homogeneous groups (agroclimatological zones, farming techniques, ownership of the field, crop mixtures, soil types etc.). The more data that are available, the more precise this classification can be. If the sample is large and the number of fields in each class is also large, it is possible to analyze the influence of the factor(s) being studied (e.g. the quantity of fertilizer used or the time of sowing). If the sample is only small, the classification should be limited to only a few classes; otherwise the analysis cannot give any significant results.
A comparison made between traditional yields, the yields achieved in the different classes, and the yields expected from the improved agricultural techniques gives a first estimate of the results of the extension program. It can also reveal the potential of that crop for the future. To analyze the results of the extension program in more depth, one can compare the results from the fields on which the extension program was followed with those from the fields that were cultivated traditionally (with perhaps some classes in between). Comparing the two extremes shows the effect of the extension package under current farming conditions. If a significant difference is found, further breakdown may indicate whether the difference can be explained by one variable (e.g. fertilizer use) or by a combination of variables (e.g. fertilizer use in combination with a specific variety and early planting).

One should be careful, however: at the level of precision likely to be achieved in farm data, it may be difficult to see a direct relation between one variable (e.g. seed variety) and the yield because there are too many other factors influencing the results (timing of agricultural operations, labour input etc.). As well, soil conditions and rainfall can vary over short distances, and the timing and quality of farming operations can vary between farmers, so the influence of a better seed variety can be overruled by other, more limiting, factors. Yields can therefore vary greatly from one field to the next. Because of this variation, when the average yield for a group of farmers is being reported, this should not be given alone, but should always be accompanied by an indication of the range or the frequency distribution.

10.3.2 Value of crops

To compare the relative advantages of different crops and to see whether one is more worthwhile for the farmer, the production achieved must be converted from the original data (in kilograms) into a monetary value, after deduction of all production costs. A price per kilogram will therefore be attributed to each crop, an important decision which can greatly influence the final conclusions about those crops. For this evaluation, the real value of the inputs and outputs to the farmer have to be used.

The prices of inputs distributed via official channels are usually easy to find out, and the prices of tools and inputs sold at the local market are usually quite stable. It is more difficult to put a price on inputs taken directly from the farmer's stock (seeds from his granary, manure) and especially on hired and exchanged labour. This must be done, however, before farm results can be calculated. (It is wise to discuss such prices with people familiar with prices in the area.) The prices of the crop are also difficult to determine: official prices are often too low, and prices at the local market fluctuate.

A good estimate of the prices at the local markets can be found by doing a long-term market survey. Local markets in West Africa are often free and open markets, the price of a product being entirely determined by the law of supply and demand. These prices therefore vary from place to place, within the year, and even within one market on the same day. A difference in price will also be found between crops sold in small quantities (e.g. a basket of about 5 kg sold by one farmer to another) and crops sold in large quantities (e.g. a sack of about 50 kg sold by a farmer to a merchant). For crops sold in large quantities, the price per kg is usually lower. This lower price should be used to value the crops, because if the farmer has a large marketable surplus, he is likely to sell much of it to a merchant.

Since prices vary over the year, the value to be used depends upon the level of detail at which the farm and price data are obtained. For an analysis of the farm income, the average price a merchant pays over the year will suffice. If the farm surveys have gathered detailed data on crops sold by the farmers and detailed market prices, several different prices can be used for the evaluation, depending on when the crop was sold. If the farmer has produced more of a certain crop than he needs, but he does not sell it, his production can be valued at the average market price to a merchant. If the food crop failed and the farmer had to purchase part of his food in small quantities, his purchases can better be valued at the higher market price asked by the farmers selling small quantities. Alternatively, one can use the price the farmers are most likely to get, which is the price just after the harvest when it is at its lowest of the year. This assumes that the rise in price during the year reflects the cost of storage and loss in the granary.

To find out the real value of a crop for one season, one has to wait till the next crop season is finished in order to know the complete price cycle. Usually, however, a long-term average price will be an adequate indicator and can be used in most evaluations. The length of the period considered when calculating the average is usually 12 months, although not necessarily from January to December. Inflation is not considered in this evaluation; that would go beyond the scope of this book.

10.3.3 Comparisons between different crops

After a monetary value has been attributed to each crop, the results of different crops can be compared. To do so, the gross margin per crop is calculated; this is the total value of the crop production (= monetary value) minus the total production cost, in cash and kind, directly attributed to that crop (= direct cost). Figure 10.1 shows how this is done.

For a crop, the gross margin alone does not say very much because many factors (e.g. the area cultivated or the labour put into its cultivation) are not the same for every crop. Crops are therefore compared according to the gross margin for the most limiting factor, which is likely to be land, labour, or capital.

- If land is the most limiting factor, the gross margin per unit of land can be calculated and the results of different crops with the same soil and water conditions can be compared;
- If land is plentiful, and labour is the most limiting factor, a presentation of the return to labour (the gross margin per day of

Crop yield (kg/ha)	X	Area under that crop (ha)	=	Total crop production (kg)
Total crop production (kg)	X	Price per kg (CFA)	=	Monetary value of that crop (CFA)
Total value of crop (CFA)	−	Direct costs for that crop (CFA)	=	Gross margin of that crop (CFA)
Gross margin of crop (CFA)	÷	Area under that crop (ha)	=	Gross margin of crop per ha (CFA/ha)
Gross margin of crop (CFA)	÷	Days of work on that crop	=	Gross margin per day of work (CFA/day)
Gross margin of crop (CFA)	÷	Number of labour units on farm	=	Gross margin per labour unit (CFA/labour unit)
Gross margin of crop (CFA)	÷	Total direct expenses for that crop (CFA)	=	Benefit/cost ratio of that crop

Figure 10.1. Relations between certain agro-economic terms

work) can be a useful indicator. For the same situation the gross margin per unit of labour can be applied.

- As a third indicator, the benefit-cost ratio for each crop can be used. This is the gross margin divided by the costs incurred to achieve it. As subsistence farmers have little capital, they may be more willing to invest in crops with a high benefit-cost ratio. However, if costs are close to zero because no inputs except seeds have been used, this ratio does not provide any useful indication. It should be used carefully.

By comparing these ratios for the different crops in a given year and comparing their development over the years, one may find some trend and may therefore be able to suggest a revised cropping pattern better adapted to the farmers, taking into account their needs for cash and food, and the risks involved. The revised cropping pattern, including input levels to be used, can then be included in the extension program.

10.4 Agricultural results per farm

As mentioned earlier, an evaluation of the separate crop results does not provide sufficient information on which to adjust project planning or programs. Cultivation practices and different crops on one farm are interwoven, with each individual deciding which crop he will put most of his effort into, what cropping pattern he is going to follow, whether he chooses for intensive or extensive cultivation etc. There are usually several decision-makers in each household. When a farm is to be analyzed as a unit, it should be realized that such a production unit is more a unit living together than a unit deciding together about farming practices.

10.4.1 Aggregation of crop results

If all the data are gathered from the same sample, a complete account can be made of the farm results for each farm in the sample. The gross margin of the various crops grown on one farm in one crop season can then be added up to obtain the gross production value of the farm.

However, this does not yet represent the real farm income, because there are a number of expenses, not directly related to a specific crop, which have not yet been deducted. These indirect costs include:

- The cost of tools;
- The salary of any permanent employee;
- All costs related to the use of animal traction: depreciation of the equipment, decrease (or increase) in value of the animals used, costs of credit, insurance, feed and care of the animals, and repair and maintenance of the equipment;
- Taxes as far as related to agricultural activities.

Some indirect costs are recurrent costs, costs which occur every year. They include: salary of a permanent employee, tools that last only one season, repair and maintenance of equipment etc. Other indirect costs are the result of an investment: the purchase of a means of production which lasts longer than one season (e.g. draught animals and equipment). Many such investments are bought on credit. The amount that the farmer has to pay to the credit institution each year covers not only repayment of the loan, but also interest on the outstanding loan for that year.

The duration of the credit is often different from the number of years during which the investment can be used. Its depreciation (yearly loss in value) must be calculated because this loss in value is a cost. Whether the investment is purchased with cash or on credit, there are many ways of calculating depreciation. It is possible to estimate the exact loss in value for each individual year, but this is not easy. A more simple approach is to assume that the loss in value is the same for each year. Then the yearly depreciation is equal to the purchase price, minus the residual value, divided by the useful life of the investment in years. This can be different from the yearly repayments, which include repayment of the loan and interest costs.
It sometimes happens that the repayment of the loan (not including the interest) is higher then the depreciation of the investment. If so, the farmer is repaying his loan and investing at the same time. This is more than just an accounting procedure; it has direct consequences for the income really available to the household.

The sum of all indirect costs can now be substracted from the total gross production value of the crops to obtain the net profit from all crops. This net profit is rarely the total income of the household. Other income will be derived from livestock, crafts, or petty trade done by one or more of its members. The net profit represents the value of the agricultural products available to the household to be consumed, stored (invested), or sold. It also represents the salary of the household for the work done in the fields.

10.4.2 Analyzing agricultural results

Some useful indicators that can be calculated at this stage will now be discussed.
If a complete farm accounting can be done for all the farms in the sample, the sample can again be divided into homogeneous groups (classes). Now the net profit of the farms in different classes can be compared or relations can be sought between net profit and farming practices within each class. Before beginning these analyses, it is wise to check whether the assumption of homogeneity within each class of farms is realistic.
When comparing different crops, ratios such as the result per unit of land or unit of labour are helpful in understanding which crop had the better result, taking the most limiting factor into consideration in the ratio.
When comparing different farms within the same class, the same ratios can be used to indicate which class of farm had the best results, again taking the most limiting factor into consideration.
If land is the most limiting factor, farms will be small because the farmer is not able to obtain any more land. The net return per unit of land makes it possible to compare the results of farms of different sizes and to evaluate which type of farm used the land resource best and so had the highest revenue per unit of land. For this analysis, one must have a detailed inventory of the key elements of the farming practices. The ratio itself is calculated in the same way as before, now using the net profit of the farm divided by the total area under cultivation.

This ratio assumes that land is a homogeneous factor. Sometimes, however, there are significant differences in the quality of the land and therefore in its potential at the level of farming practised in the area. In such cases it is not possible to use the ratio of net profit per unit of land. If labour is the most limiting factor and the farmer is not able to cultivate all the land he has available, the net profit per unit of labour or man-day of work can be used as a means of comparing farm results. If the exact use of labour for the cultivation of the crops is not known, the net profit per unit of labour can replace the net profit per man-day of work. The net profit per unit of labour is the net profit from the farm divided by the number of labour units that the household has available for agricultural work. The net profit per man-day is the net profit from the farm divided by the number of man-days devoted to farming activities. In the analysis of the farms, the farmer with the highest return to labour has been the most efficient. It is quite possible that such a high return to labour coincides with a very low return to land, which is not important here because land is not a limiting factor.

The quality of land can vary even within a field

If capital is the most limiting factor, the benefit-cost ratio can be used again. It is the net profit of the farm divided by the total costs (direct and indirect) of that farm. In farms where very few inputs are used, this ratio has to be used with care, as was explained in 10.3.3.

The analyses presented above are purely economic and are the most simple to perform. They have identified the farm type with the best results, but the analysis is not yet finished. It has revealed that most farmers do not farm in the way that would bring the highest benefit. An analysis of the background information gathered about the farmer might help to explain why he acts in the way he does. All kinds of reasons, not easy to quantify, can play a role: social obligations, dislike of certain foods, problems in marketing. An analysis of these factors might point out the need to do some additional surveys or to include more specific questions in the surveys of the following crop season.

Drawing conclusions about a more efficient farming system that is also acceptable to the farmer is not a thing that can be done in a hurry. Nevertheless preliminary conclusions can help the extension agency to design a program which is already better adapted to the situation of the farmer.

10.5 Income from livestock

Few farms derive all their income from crops; most have some livestock as well. Raising sheep, goats, and poultry brings a small but steady income to many farms. More importantly perhaps, livestock represents a form of savings, readily available when needed, since livestock can always be sold. In prosperous times, benefits from crops are invested in cattle and the new-born are kept in the herd. If the crop fails one year, the farmer can draw from his 'savings' and sell an animal once in a while to buy cereals. Chickens and sometimes animals, are sacrificed during rituals or slaughtered for festivals. They also feature prominently as gifts or as exchanges required by social custom. Traditionally, little care is given to poultry and livestock, and death rates are high. In some farming systems, livestock and crop production are integrated: through the cultivation of fodder crops or the systematic use of crop by-products for feed. In other systems, livestock and crop production are done by different groups of people.
The evaluation of farm results should analyze the main types of livestock, the scale at which they are raised, and the benefit derived from

them. The level of detail needed to evaluate this activity depends in part on whether the project includes, or plans to include, a livestock component.
For integrated farms, it is customary to calculate the net farm profit of crops and livestock as one entity in the total farm income. For farms where crops are not linked to livestock, one takes the net profit of crops and places livestock separately in the total household income.

It is difficult to evaluate the income from livestock because most of it is obtained in kind: through the added value of the animals, births within the herd, gifts, and so on. It is equally difficult to estimate what an animal is worth because that depends on its weight, health, and age, and because at most local markets there is little trade in livestock except poultry. In a monitoring program, however, data can be gathered on the amount of cash involved in livestock keeping and this information, combined with a regular inventory of the size of the herd, can provide rough estimates of the income generated from livestock.

10.6 Total household income

Even though farming and livestock may be the main sources of income of the households studied, most of these households will also earn some income from other sources (e.g. trade, handicrafts, or paid employment of one of their members). If the income from these sources has been surveyed, the net income derived from them can be added to the net profit of the farm to obtain the total income of the household. This does not specify the distribution of income among its members, but only the total amount available to them.
The total income is an important indicator of the household's well-being, but it does not represent the amount (in cash and in kind) available to its members to take care of their needs. Part of this income is invested in means of production (plough, seeder) so that the income the household can really use is often less, unless some livestock is sold (thereby reducing the value of the herd) to buy some necessary food or clothing.
To estimate whether the total income is satisfactory, the most basic

needs, at least, have to be quantified. Since all the expenses involved in crop production have already been deducted, the total farm income covers the following expenses:

- Food consumption of the household; this will be discussed in the next section;
- Cash expenses for such goods as cloth, household implements, housing, bicycle repairs, etc. Through informal inquiries, these expenditures can be estimated and an average budget established. If more precise data are needed, a farm budget survey can be organized;
- Health and education;
- Taxes;
- Social obligations; these costs are not easy to estimate but they can be quite high;
- Investment in income-generating activities; to keep his enterprise going, the farmer may need to make regular investments in agricultural equipment, in stock for trading, and in equipment for other income-producing activities. Even when the investments are financed by an agricultural credit organisation, the farmer usually has to make a down-payment.

If the expenses listed above are less than the total net income, the farmer has funds available for less urgent expenses and for additional investments such as those in livestock.

10.7 Food sufficiency

To evaluate the net income of a subsistence farm, one must know whether the crops produced cover what the household needs to feed itself. The total food requirement of the household is obtained by multiplying the estimated needs per person by the total number of people who have to live from the farm production. The most common practice is to work with food requirement standards of FAO, World Bank, or other sources. The evaluation should clearly state which minimum requirement has been selected as standard. In a traditional farming system in which cash crops play only a minimal role, the caloric value of the production could be compared with the caloric requirements of the household. This assumes that if the caloric requirements are satisfied, the protein and vitamin requirements are satisfied as well.

Many projects, however, introduce new, relatively expensive technology for food crops, in combination with a cash crop, so that the farmer can buy the necessary inputs. In such projects, two aspects of food sufficiency need to be evaluated.

First, the comparison between food requirements and food production show the food sufficiency in kind. If no adequate official marketing channel exists, this is an important finding to predict the extra supply or demand that the project is likely to create on the local market. A shortage of food crops can lead to a decrease in the area given to cash crops in the following year, because a farmer will primarily try to satisfy the food requirements of his household and to store some food for bad times.

Second, to evaluate the food sufficiency in monetary terms, the monetary value of the total household income can be compared with the monetary value of the household's food requirements. This shows whether the farm meets the minimum economic standards for survival.

The range of monetary surplus and the frequencies of households per class of surplus can then be calculated, as well as the percentage of households who have not produced enough to cover their food needs. If need be, the data can be regrouped according to the strata of the sample, level of technology, or any other aspect used to explain the difference in monetary surplus.

The results of such an analysis, which will indicate some of the causes of insufficient production, will provide the extension service with the information it needs to revise and strengthen its programs.

11 RESULTS AT PROJECT LEVEL

11.1 Farm types

Once the total farm income has been calculated for all the farms in the sample for which complete sets of data are available, the average income can be calculated for the whole sample and for the various sub-groups within the sample. When reporting the results, it will be helpful to describe the economics of specific groups of farms. These groups can be the stratifications used for selecting the sample in the first place, or they can be those farms that have been found to follow a common pattern of cultivation. The stratifications could distinguish modern and traditional farms, or farms with a high man/land ratio and farms with a low man/land ratio. It is also possible to describe the economics of the least successful farmers to indicate their problems, and the economics of the most successful farmers to indicate the potential of farming under that level of technology. These descriptions should include: cost structure, gross margin per crop, farm income, etc.
Calculations should be done for each agro-climatological zone separately, and only for groups that are somewhat homogeneous. This provides the kind of information that could lead to the addition of a new survey during the next crop season.

11.2 Net profit of a theoretical farm

Chapter 10 explained how to calculate the total income of each farm and the average income of a group of farms when information is available on every activity of each farm. To do this, all the data must be gathered from one single sample. Often, however, different types of data are gathered from different samples: the cropping pattern and size of household from one sample of farms, the production of crops from several different samples of fields, and the income from off-farm activities from yet another sample. The total income of any of the farms in these samples cannot be calculated. If, however, the samples are all selected in the same way and are thus representative of the same total population, the total income of the average farm representing a theoretical farm from that population can be estimated. This is done by following the steps described in Chapter 10, using the average data from each sample as the data for the theoretical farm. This will be explained by an example.

Example of estimating average crop results

For surveys conducted in a village, the samples were stratified into two groups: farmers with and without animal traction. The surveys done were the following:

- One survey on a sample of farms to gather information on:
  - Number of fields and type of crop cultivated on each field;
  - Indirect costs;
  - Size of the household, and number, sex, and age class of the workers.
- Other surveys on samples of fields, one sample for each type of crop. These samples could be sub-samples from the first sample, but can also be completely different samples. Each survey gathered the following information:
  - Yield, production, and area of each field;
  - Direct costs.

The average results of the surveys are shown in Table 11.1. Using these averages as if they were the data on one theoretical farm, representing one group of farmers (those with animal traction or those without), the average gross margin of each crop of that group can now be calculated. The prices used to value each crop are the

Table 11.1. Average crop results for two groups of farms

Crop / Average value of variable	Farmers with animal traction					Farmers without animal traction			
	Sorghum	Millet	Peanut	Cowpea	Cotton	Sorghum	Millet	Peanut	Cotton
Production per field (kg)	700	460	510	170	940	770	490	690	470
Price per kg produce (CFA)	62	65	81	79	54	62	65	81	54
Value of production (CFA)	43,400	29,900	41,310	13,430	50,760	47,740	27,950	39,690	25,380
Direct costs per field (CFA)	4,900	420	3,200	800	9,700	4,900	370	3,500	4,800
Gross margin per field (CFA)	38,500	29,480	38,110	12,630	41,060	42,840	27,580	36,190	20,580
Number of fields per farm	2.1	0.7	1.2	0.3	0.8	1.5	1.1	0.9	0.3
Gross margin per crop (CFA)	80,850	20,636	45,732	3,789	32,848	64,260	30,338	32,571	6,174
Gross production value per farm (CFA)	183,855					133,343			
Indirect costs (CFA)	18,900					3,700			
Net profit per farm (CFA)	164,955					129,643			
Area cultivated per farm (ha)	4.17					3.2			
Net profit per ha (CFA)	39,558					40,513			
Number of labour units per farm	4.2					3.7			
Net profit per unit of labour (CFA)	39,275					35,039			
Ratio net profit total (= direct + indirect) cost	4.35					7.51			

average prices at the local market that year (Chapter 10 Section 3.2).

For farmers with animal traction, the average gross margin from sorghum is calculated as follows:

- Value of production per field (CFA) = average production per field (kg) × price per kg = 700 × 62 = 43,400 CFA;
- Gross margin per field (CFA) = value of production of field (CFA) - average direct cost per field (CFA) = 43,400 - 4,900 = 38,500 CFA;
- Gross margin per crop (CFA) = gross margin per field (CFA) × average number of fields for that crop = 38,500 × 2.1 = 80,850 CFA.

After the same calculations have been done for each crop, the average gross production value of the theoretical farm can be calculated.

- Gross production value of the farm (CFA) = sum of gross margins per crop (CFA) = 183,855 CFA;
- Net profit per farm (CFA) = gross production value of the farm (CFA) - indirect cost (CFA) = 183,855 - 18,900 = 164,955 CFA.

The results are calculated in the same way for the farmers without animal traction. From these results, various ratios can be calculated as described in Chapter 10 Section 4.2.

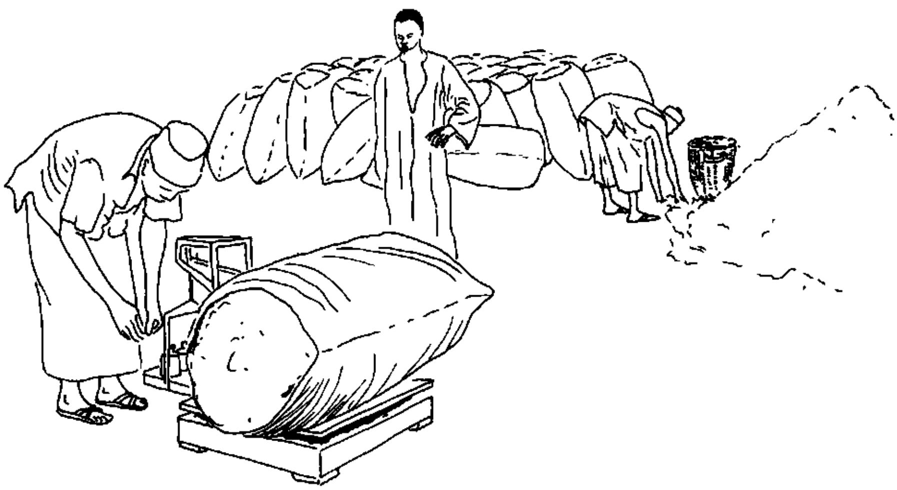
11.3 Extrapolation of results

When data have been gathered on one or more samples, each representing the entire project population, the average results of the sample(s) can be used as the average results of the population. The various productions, incomes, and ratios calculated for the sample(s) can be extrapolated to apply to the entire project population.

For example: If the total number of farms in the project is known, it is possible to extrapolate the total production in the project by multiplying the average farm results by the total number of farms. This can be done with the average farm results of the entire sample, but also with the average farm results of groups of farms (or strata) if the total number of farms in those groups is known. In the second case, the average of each group has to be multiplied by the number of farms in that group, to ensure that each farm in the sample is represented equally.

The population results can be useful to the project managers for two purposes:

- To present the results of the project to the planners and development agencies involved, and;
- To plan for the future needs of the project population.



Project should know what surplus is available for sale

Example of extrapolation

On a sample representative of all the farms in the project area, it has been established that the average sorghum production was 2,150 kg per farm. There are 2,600 farms in the area. Thus the population produced a total of 2,150 × 2,600 = 5,590,000 kg or 5,590 tons of sorghum. If the average number of consumers per farm is known, the total quantity of sorghum that the population needs to feed itself can be calculated (using estimated requirements per person as described in Chapter 10 Section 7). Suppose there are 8.2 consumers per farm and it has been decided that the standard sorghum requirement is 240 kg per person per year; the total population then needs 240 × 8.2 × 2,600 = 5,116,800 kg = 5,117 tons of sorghum. The surplus production of sorghum that year is 5,590 - 5,117 = 473 tons.

The project should know what surplus is available and might be sold, or whether a food shortage is occurring. If a surplus occurs and managers can estimate the stock available for marketing, they can review the various marketing channels available for the different crops (local market, merchants, government marketing board), and plan for improvements in storage and transport as necessary.

11.4 Estimating future needs

Extrapolation can also be done for inputs applied in the sample. The use of fertilizer, insecticides, ploughs for animal traction, credit facilities, and so on can all be extrapolated, so that supplies and distribution channels can be prepared for the next crop season.
Another extrapolation can be made if a trend can be observed in the data. Such a trend can only be seen after several years of monitoring. If a trend occurs and it is clear that it is not due to random factors, that trend can be extrapolated to the future.

For example: It has been observed that the production of cereals has been increasing over the last few years, but at a rate that varies from year to year. An estimate of the cereal production over the next few years can be made by extrapolating the trend already existing. In this way, likely surpluses or shortages can be identified and management can plan early to adapt its strategy to deal with them.

11.5 Simulation

If data have been gathered about all the aspects of agricultural production - whether on one sample or on different ones - the data can be used to simulate a farm. This means putting a hypothetical farm together with the information obtained during monitoring. This hypothetical farm is helpful in finding an economically better farming system. After defining the hypothetical farm, it is possible to change one variable at a time and analyze what the result will be.

For example: What will happen if the cereals fail or if the price of a certain crop goes up?

This is a very useful exercise to obtain a better understanding of how the farming practices are interrelated. If the monitoring unit has access to a computer, such a simulation can be done more systematically, changing more variables at the same time.

12 REPORTING

12.1 General

The reports of a monitoring and evaluation unit present the results and the conclusions derived from the analyses of farm data. The reports will be read not only by the managers of the project but also by field personnel and possibly by people outside the project. They should therefore be presented in such a way that they can be understood by people who are not involved in the project.

12.1.1 Presentation

The information in a report should be easy to find. People looking for a specific item of information are not going to read the entire report. A report should be clear and systematic, with many titles and subtitles to indicate where information can be found. Reports are easier to read if the numerical information is presented in simple tables or graphs. Graphs help the reader notice trends in the data and are a good way of presenting figures that are not very precise. They also make the report more attractive to the reader. A brief summary of the report's main findings will be helpful to the readers.

12.1.2 Types of reports

A monitoring and evaluation unit will usually be required to produce two types of reports:

- Preliminary reports;
- Evaluation reports.

A preliminary report is a first rapid report of the information obtained in the field, sometimes still incomplete. Its main priority is speed in reporting, so concessions can be made in the type and quality of the information presented. Examples of such reports are regular monthly reports on agricultural activities or reports about the season's agricultural production, written early in the harvest period.
An evaluation report is a more thorough study of the process being evaluated. Such a report requires more time to prepare and includes a detailed analysis of the data, something not possible in preliminary reports. Examples of such reports are the final report on an agricultural season, or a report analyzing trends over several agricultural seasons.

12.2 Evaluation reports

Preliminary and evaluation reports have different contents and different presentations, although there is much overlap. Since the evaluation report is the most complete of the two, it will first be discussed.

12.2.1 Readers

The main readers of an evaluation report are the members of the project who helped to delimit the work of the monitoring and evaluation unit, but the report is also likely to be used to present the project achievements to funding agencies and potential donors. The report also serves to make the knowledge accumulated within the project available to others working under similar conditions.
Before he begins to write a report, the unit director, who will usually be called upon to produce the unit's reports, should find out what kind

of people are likely to read it in addition to project members, so that he can adapt the contents to their needs. An internal report can assume that its readers have some knowledge of the project, but a report intended for both internal and external use must include a detailed description of the project.

12.2.2 Contents

The contents of a report and the way the information is presented will, of course, be in accordance with the requests made during the early consultations with the users of the information. Often, however, instructions about presentation do not cover more than requests for certain types of graphs and indications of deadlines. The unit director is therefore usually free to decide how the information will be presented. Suggestions for the outline of a report, covering the subjects usually treated and arranged in an order logical for the reader, are given below.

Section 1: A thorough description of the project area

- Location, accessibility of area;
- Rainfall pattern, agricultural season, soils;
- Population, household size, labour force;
- Marketing channels;
- Land tenure;
- Traditional farming systems.

This description is needed even if the results are only for internal use. It shows the basis on which the project is taking place. These details can help explain the results achieved by the farmers or by the project. Besides, not all project personnel are aware of all these facts.

Section 2: A description of the project

- Goals of the project;
- Means available to achieve the goals;
- Detailed description of agricultural and other innovations being introduced by the project.

This may seem superfluous if the report is only for internal use, but goals are often modified during project implementation and the agricultural extension program is frequently adjusted. Many of these changes are made in an informal way and are not always recorded. It is sometimes difficult to find out what the project is really promoting. A description of the project is therefore necessary to show what starting point the unit took in its analysis of project achievements.

Section 3: Methods of data gathering

- Methods used to select the samples;
- Sample size;
- Survey program;
- Validity checks;
- Training of the staff of the unit.

This section describes how the data on which the evaluation is based were gathered, and gives an indication of the reliability of the data. The readers who will use the results should know how the data were gathered, how the samples were selected, and what steps were taken to ensure the validity of the data. They can then decide for themselves how they will use the results. If it is clear that the work was technically as correct as possible under the circumstances and was performed with care, the credibility of the information presented later is enhanced. To make a detailed feedback from the users possible, the report can present, in an annex, copies of the survey forms that were used, to show exactly what kind of data were gathered.

- Prices used in the evaluation;
- Crop results;
- Farm results;
- Other income;
- Extrapolation of results to the future.

This section presents the data gathered in the way requested by the users. The groups and classes that are distinguished have been selected in accordance with the users' requirements. The references against which the results are compared were identified by the users and must be explained. One of these references can be a goal of the project, but there are many others, such as a standard food requirement to evaluate whether the farmer harvested enough to feed his household, or a standard labour input to evaluate whether the farmer could cultivate a crop more intensively.

Even in a first report, it can be useful to extrapolate the data to the future, stating clearly the limitations of such a forecast. It shows the users what kind of forecasts can be made in the following years, so they will know better what to request in future.

Section 5: Project results

- The total results of the project.

A project's monitoring and evaluation unit will seldom be asked for an evaluation of the entire project because this is an evaluation that should rightly be done by an independent agency. Nevertheless, the unit can supply the basic data for such an evaluation, from which managers can see whether the project is indeed achieving its goals and whether they should intervene to adapt project strategies.

Section 6: Conclusions

This section groups in a logical way the important conclusions mentioned throughout the report, combining conclusions pertaining to the same topic.

Section 7: Recommendations

Here the monitoring and evaluation unit suggests possible ways in which the users could adapt their line of action to the situation in the field. It may also propose further research.

Section 8: Summary

This is often combined with the conclusions, but it is useful to present it as a separate section which can be read quickly. If presented as a separate section, it should also contain a summary of the most important conclusions and recommendations. Sometimes a translation of the summary may be needed. Summaries are occasionally placed at the beginning of a report.

12.3 Preliminary reports

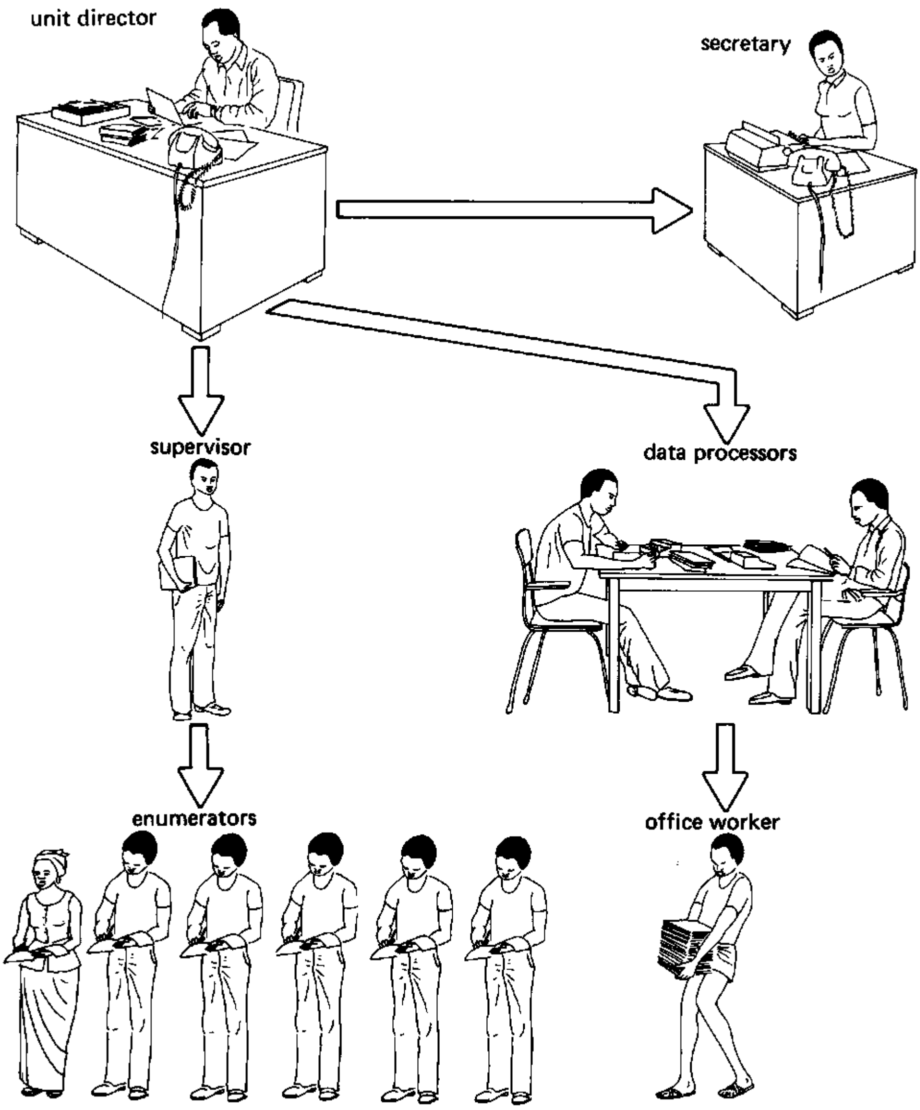
Preliminary reports present the first findings that become available while the processes of data gathering and analysis are still going on. This kind of report has to be issued as early as possible, so concessions can be made in its contents and presentation. Although preliminary reports are far more restricted in scope than evaluation reports, they can follow the same outline, but with the general sections kept short since they are meant only for internal use. If the same outline as in the evaluation report is followed, the users of the information will become familiar with the manner of presentation and will thus have easier and more rapid access to the information they require. Preliminary reports are the major incentive for feedback from the users, and are therefore essential in the monitoring and evaluation process.

The simplest form of preliminary report is the regular presentation of the results from on-going tabulations, in the form of tables and graphs.

As always, the headings of these tables and graphs should give all the information needed for anyone to understand them without having to read the text.

Different reports on the same groups of farms should present the information uniformly and in the same order in each report, so that the users can easily compare the results from successive reports by placing the tables or graphs side by side.

Most of the data presented in preliminary reports will be average farm results and their distribution. Extrapolations to project level and to the future can bring useful feedback from the users.



Staff of a monitoring and evaluation unit

13 STAFF OF A MONITORING AND EVALUATION UNIT

A monitoring unit is likely to employ four categories of staff:
- A unit director;
- Field supervisors;
- Enumerators;
- Data processors;

and possibly also its own secretary or office worker.

13.1 Unit director

13.1.1 Task

The director of the monitoring unit is in charge of organizing the unit's work program. His first task is to acquire a thorough understanding of the socio-economic conditions in the project area. He then has to find out who will use the data and what kind of data they need. Subsequently, he has to translate these needs into a detailed, workable program of data gathering, processing, and analyses.
Before such a program gets under way, the director should make sure that it is coordinated with the programs of other sections of the project, thereby avoiding duplication of effort and incompatibilities in data. He should keep informed of modifications in, and additions to, the extension and other programs. He has to keep abreast of new developments in the villages not only through the reports of his field staff but through reports from other sections of the project and by visiting the villages himself.

The director is responsible for reporting the results of the unit to the users so that they can incorporate these results in their future planning.
The director will also organize and partly conduct the training of the staff.

13.1.2 Level of education

Much is asked of the unit director. He is at the same time an administrator, a team leader, and a survey specialist. Even if he is assisted by a statistician, he should himself be able to design the broad lines of the program of farm surveys and analyses. The most appropriate person for the job would be an university-trained agro-economist, although it is emphasized that the main qualities required are common sense, the ability to organize a program and follow it through, and a knowledge of local rural conditions.
A knowledge of statistical theory is less important because most of the analyses to be done are quite simple. If the unit is requested to gather a large body of data, the director may need the assistance of a technician knowledgeable of basic statistics and possessing some practical experience in data processing and analysis. The technician can help organize the work of the data processors and set up the tabulations to be done in the field, but he need not be with the unit all year around.

13.2 Field supervisor

13.2.1 Task

A field supervisor is in charge of a group of enumerators. His task is:
- To check the work of the enumerators. He does so by visiting his enumerators regularly to see whether they are working to schedule and to check the validity of the data they have gathered;
- To coordinate the work of the enumerators. If an enumerator meets a problem that he cannot solve, he talks it over with the supervisor during one of the supervisor's visits and together they look for a solution; the supervisor will then inform his other enumerators of

this solution so that everybody adopts the same methodology. Part of the supervisor's coordinating task is to arrange help for an enumerator who has an excessive work load in a certain period so that he does not fall too far behind the others;

- To guide the enumerators in their work. Especially at the beginning of a survey, its objectives have to be explained to the enumerators in great detail. Often this explanation proves inadequate and enumerators will ask for further explanations or more details later on. The supervisor has to reserve much of his time for explanations, going through the survey forms with the enumerators and telling them how the data will be processed. If an enumerator has made an error in filling in a form, the supervisor should not just correct the error, but should also explain why it was wrong;
- To group the data. To check the validity of the data and to speed up their subsequent analysis, the supervisor does a first grouping of the data, simply transferring the data about individual fields or farmers from the survey forms to recapitulative forms at the crop or village level.

13.2.2 Level of education

A supervisor should be able to guide the enumerators so he may need a higher level of education and certainly more field experience. About eight to ten years of schooling should be sufficient. Any level higher than a high school diploma would probably not improve his performance with the farmers. He should have previous field experience in data gathering and be able to establish good relationships with the farmers; for this, he needs to speak the local language. One of the main aspects of his job is checking the validity of the data gathered, which can only be done well by someone who is familiar with local village life and traditional farming techniques.

13.2.3 Number of supervisors

The number of supervisors required depends on the number of enumerators.

The number of enumerators controlled by one supervisor depends in part on transport facilities. But even if travel to the villages is easy, the supervisor is not likely to work well with more than about six enumerators, since he must also visit the unit's office regularly. Most enumerators need to be visited at least every two to three weeks, depending upon their previous experience. Even a very small unit should have one supervisor, so that field visits can still be done on schedule when the unit director is busy with administrative tasks and meetings.

13.2.4 Working conditions

Ideally, the supervisor should live in a village from which he can easily visit his enumerators, even during the wet season. From there, he will pay regular visits to the office to bring the completed questionnaires and recapitulative forms, pick up supplies, and report to the director on the situation in the field and the work accomplished by each enumerator.

13.3 Enumerators

13.3.1 Task

The enumerators gather the data required by the unit. The data are gathered from farmers who usually cannot read or write and therefore do not keep records of their activities. The enumerators will be called upon to conduct single-visit surveys, but their main task will be to survey continuous processes by repeated interviews with the farmers and by observations of farming activities. They should therefore be familiar with local conditions and should live in the area being surveyed.

13.3.2 Level of education

Human qualities are more important criteria for the selection of enumerators than education. A high school diploma (bacalaureat in Francophone countries, fourth form in Anglophone countries) can be more of a handicap than an asset. A level of six to eight years of schooling is often better, for the enumerators are then still willing to live in rural areas and do not consider themselves too different from the farmers. Previous experience in survey work can be an asset, but is not essential.

Selecting and training the enumerators are crucial steps, since the validity of the data they gather will determine the validity of the unit's entire work. An enumerator should be able to work on his own, be honest enough that he neither invents information nor writes down 'what the farmer really meant to say' (although he did not say it), and he should be able to establish good relationships with the farmers.

Basic requirements for enumerators are likely to include:

- An open, friendly personality, which will put the farmers at ease. A good enumerator should be genuinely interested in the life of the village and in the farmers' activities;
- Fluency in the language of the farmers and familiarity with their social organization and farming methods;
- Ability to work accurately even when alone;
- Literacy and a basic knowledge of mathematics;
- Willingness to live in the village he will survey. An older, married enumerator will have a better chance of being accepted by the farmers.

13.3.3 Work load per enumerator

When deciding on a reasonable work load for an enumerator, it is safer to be a bit pessimistic because the quality of the survey soon declines if his interviews or observations are done in a hurry or out of schedule. Visits to the field to observe crop performance and the farmers' activities or to estimate crop yields can take up much of his time. The actual interviews with the farmers and filling in the questionnaires and tables do not take long, but finding the farmer to be interviewed and talking informally with him before and after the interview can be very time-consuming. Travel time should be taken into account because the houses and fields may be scattered over a wide area. Often it is safer not to plan on more than four or five interviews a day.

The work program of the enumerators has a better chance of being correctly followed if it forms a well-established routine. In a repeated survey, the same farmers should be visited on the same day of the week so that neither the farmer nor the enumerator forgets the appointment. The visits can, of course, be postponed for a day to avoid working on a holiday. It is sensible to plan one working day a week with no interviews, during which the enumerator can do his 'office work' and catch up on interviews or other work that has been delayed. If necessary, this day can also be used to do some additional single-visit surveys.

The entire work program can be described in detail in the 'Enumerator's Manual' (Chapter 6 Section 5.2) written by the unit director. This manual should also give a calendar of activities for the entire crop season.

13.3.4 Number of enumerators

The number of enumerators needed depends upon the size of the samples and the program of surveys. In deciding how many enumerators are needed, the director should, of course, observe the social laws of the country concerning days off and paid vacations and should not expect enumerators to work on official holidays and on weekends if the project personnel does not either. It should be planned that enumerators' vacations are taken outside of the cropping season.

13.3.5 Female enumerators

Women are actively involved in agricultural production and other activities likely to be covered by farm surveys, so some information is better obtained directly from them rather than from their husbands.

> For example: Women can be asked to supply data on labour, the production of their personal fields, handicrafts, petty trade at the market, and the food consumption of the household. In many cultures, a male interviewer is restricted in his behaviour towards women. He

cannot enter a compound when the husband is absent, can only talk to a woman in the husband's presence, and even then it is often the husband who will answer for her.

If data on women's activities are desired, some female enumerators should be employed by the unit if at all possible. They should fulfil the same educational and personality requirements as the other enumerators.
Female enumerators are particularly needed in a program of long-term monitoring, because the focus of data gathering is likely to shift from purely agro-economic data to more socio-economic data (e.g. the distribution of labour and income within the household, off-farm activities). In the beginning, female enumerators may not seem necessary but in the long run they are essential for gathering valid data.

13.4 Data processors

13.4.1 Task

Data processors check and tabulate the data and perform basic analyses. They work in a central office under the supervision of the unit director.

13.4.2 Level of education

The educational level of data processors should be as high as possible. This will, in all probability, be eight to ten years of schooling, with some additional training in mathematics. The most important requirement for data processors is accuracy because errors in their work can lead to errors in the analysis and therefore to incorrect conclusions.

Normally, data processors do not go to the field, but their participation in training sessions for enumerators will help them establish a better contact with the field workers. On-the-job training will improve their competence.

13.4.3 Number of data processors

The number of data processors needed depends on how much the supervisors participate in processing the data (on recapitulative forms) and on the size of the data base to be handled. The type of analyses desired and the medium of analysis (computer or otherwise) do not influence the number of data processors required as much as one would think. The preparatory phase of the analysis (checking the data and transferring the information onto recapitulative forms) takes a great deal of time and has to be done whether the data are processed by hand or by computer. As a rule of thumb, one data processor cannot usually handle more than the data gathered by five enumerators.

13.4.4 Working conditions

The data processors will work at the office. It is useful if each processor specialises in the tabulation and analysis of certain sets of data. They are then more likely to notice errors because of their familiarity with those parts of the surveys. Such specializitation also allows data processing to proceed more rapidly. Who is doing what tabulation and analysis should always be clear, so that in case of delays the person responsible can be identified.
The data processors are also responsible for filing the data. For this, a systematic approach is essential to avoid data getting lost, something that happens more often than one would expect.

13.5 Office worker

13.5.1 Task

When the monitoring unit is part of a larger project, it is often thought that the unit's typing and duplicating can be done by the project's office workers. Usually, however, it is better for a monitoring unit to employ its own full-time office worker, as there is much typing and duplicating to do: questionnaires, reports, manuals, recapitulative forms, and so on.

Much of the value of a survey is determined by the speed with which the data are made available to the users, so good typing facilities are essential. If no typist can be employed by the monitoring unit, there should be a guarantee of sufficient typing capacity in the project.

The duplication and distribution of the questionnaires and reports are likely to constitute a bottleneck. If duplicating is done in the project's duplication office, one of the data processors should be responsible for making sure that duplication is done as scheduled and that the forms are sent to the field supervisors for distribution to the enumerators.
The office worker (or a data processor if the unit does not employ an office worker) should also be made responsible for keeping track of the forms returned by the supervisors.

13.6 The hiring process

The selection of the enumerators and supervisors is an important step which cannot be rushed through; the quality of the surveys depends greatly upon the quality of these employees. Their personality is as important as their knowledge, so the hiring process should include a trial period during which the candidates can be evaluated not only for their knowledge and accuracy but also for their behaviour while conducting test interviews with some farmers. This is not time wasted! Personnel turnover in the field during surveys is highly disruptive and delays the completion of the surveys.

When advertising for recruitment, efforts should be made to reach potential candidates in the area to be surveyed, either through radio broadcasts or by informing local authorities. A first selection can be made on the basis of a written test and an interview, to check on the basic competence of the candidates (vocabulary, mathematics, knowledge of the local language etc.) and on their ability to follow instructions carefully and to perform detailed work. A first training session should take place before the final selection is made, as it provides good opportunities to check on the candidates' ability to learn and to follow

instructions, and on their ability to understand the objectives of the surveys.
The data processors can be selected through written and oral tests to check their knowledge and particularly their accuracy. They can always be trained to perform calculations they are not familiar with, but if they are careless when calculating or unable to keep a file of questionnaires in decent order, they will never make good data processors.

13.7 Division of responsiblity

There should be a clear division of responsibility among the staff of the unit. For each task that needs to be done, one individual should be made responsible. Being responsible does not mean that the person must always do the task himself; it means that he must see that the task is done as planned and on time.

Often a monitoring unit has a centralized structure, with the unit director ultimately responsible for every task. Such a hierarchical structure is shown in the drawing at the beginning of this chapter. This structure does not mean that the director himself has to give the order before anything gets done. Each staff member should have a work schedule listing all routine tasks that are to be done without further instructions.

13.8 Staff performance

As in any team, the performance of the unit's staff is greatly influenced by the attitude of the unit director. No one works carefully for long periods of time if there is no evidence that the work is of any use. The unit director should therefore see that the reports of the unit are distributed to the staff and discussed with them. If possible, he should provide evidence that project management is aware of the conclusions of the reports and incorporates them in the decision-making process. Such evidence greatly encourages the staff to keep up their good work.

The unit director should also arrange that some internal promotion be possible, from enumerator to field supervisor, for instance, or from the monitoring unit to a higher rank in another division of the project.

In some projects, rewards for good work are effective incentives: a monthly cash bonus, for instance, to enumerators who have had only a small percentage of their questionnaires rejected during validity checks, and to data processors who have processed a file accurately and on schedule. Bonus systems, however, are not accepted in every country.

14 STAFF TRAINING

Whether they are already experienced in data gathering or not, the enumerators and supervisors will need to be trained to observe the activities of the farmers and notice those facts important enough to be recorded. They must know and thoroughly understand the purpose of each survey and its place within the survey program. They must be able to establish good relationships with the farmers, to ask questions in a way that does not influence the answers, to listen to the farmers, and to be alert for any wrong or incomplete answers.

The data processors need to be trained in the checking, tabulation, and analysis of the data, and taught to work carefully, always checking that their calculations are correct.

Obviously, it is not possible to design one training program that could meet the needs of all the staff. Even so, a general training session for everyone should be given before the surveys begin. This session should focus on the work of the enumerators, because the supervisors and data processors should be familiar with the enumerators' work if they are to do their own jobs correctly. After this general session, the supervisors and data processors can receive more specific in-service training.

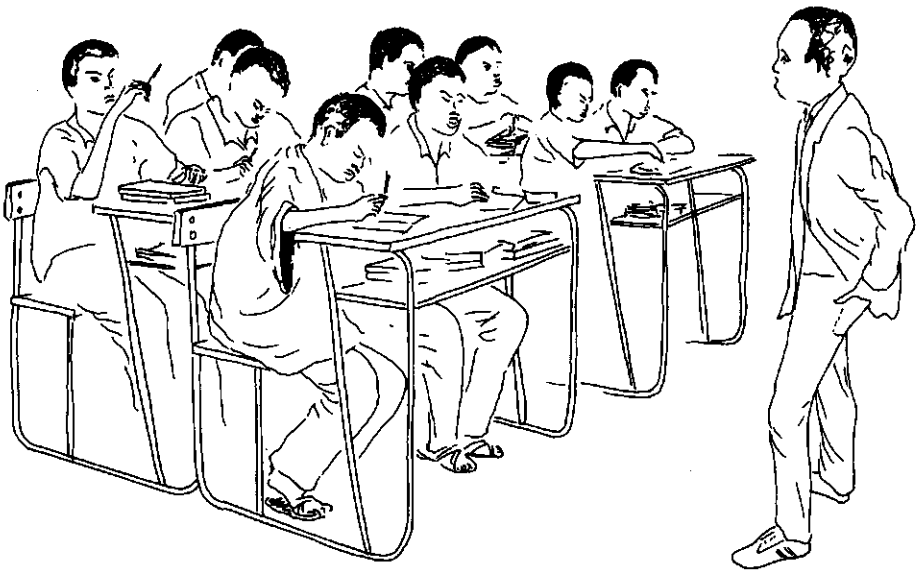
14.1 First training session for the entire staff

The first training session - to be organized before data gathering begins - might last two to three weeks. Its aim is to review - and systematically reinforce - the basic knowledge required by the enumer-

ators to perform their tasks. As text book for the training session, the enumerators could be issued copies of *Introduction to Farm Surveys*, which has been especially written for this purpose, or, alternatively, publications of *Institut Africain pour le Développement Economique et Social* (INADES) and the French Ministry of Cooperation.
The training session should alternate classroom lectures with practical exercises.

> For example: If the enumerators are required to perform crop cuttings for yield estimates, they should first be taught in the classroom what crop cuttings and yield estimates are and what method will be used to obtain them. This is then demonstrated in the field. Afterwards, the enumerators, working in small groups, practise placing test plots for yield estimates, using the equipment they will use in their later work.

During the training session, the enumerators should be issued with the equipment they will need and be shown how to use and maintain it.



Formal training before the surveys begin

The enumerators should be taught the techniques of observing and interviewing and of filling in the questionnaires. They can then practise among themselves in small groups in the classroom. Once they are sufficiently familiar with the techniques, they can be sent in groups of two or three to interview some farmers. The farmers chosen for test interviews should preferably not be included in the sample that will later be surveyed and not even, if possible, in a village that will be included in the sample.
As a final exercise in the training session, the enumerators can be asked to conduct a simple survey in a test village from beginning to end: designing the questionnaires, interviewing a few farmers, tabulating the results, and writing a simple report. Here, too, they should work in small groups so that they can help one another, and once again the test village should preferably not form part of the later sample. Afterwards, the supervisors and the unit director should go over the 'survey' with the enumerators, discussing its strong and weak points.

14.2 On-the-job training of enumerators

One single training session before data gathering begins is not sufficient. Even experienced enumerators who have mastered the interviewing and observation techniques need additional on-the-job training during data gathering. For this purpose, frequent one-day sessions (e.g. once a month) can be held in a centrally-located village for groups of enumerators working in the same area.
The goals of these training days are:
- To discuss the situation in each village, the agricultural tasks that are being done, the expectations for the crop season, and any unusual events that took place since the last session;
- To discuss the survey program as it is being implemented. The enumerators should mention their problems in data gathering, especially those that were not foreseen in the Enumerators' Manual. Together the enumerators and their supervisors can try to find solutions, which increases the enumerators feeling of involvement in the survey program;
- To review specific techniques in surveying before they are applied.

- To improve the enumerators' understanding of survey work and increase their interest in it by explaining something of the theory of basic statistics and agricultural economics;
- To provide some training in sociology. The enumerators should understand that the way they approach the farmer greatly determines the validity of the data gathered. Enumerators who have gone to secondary school and have lived in a town for some years may have forgotten much of the reality of a farmer's life. Frequently, a person with some schooling believes that the farmers' ways are old-fashioned and undesirable. Providing some basic training in sociology can obviate much of this prejudice. The enumerators should be made aware of the social structure among the people who will be surveyed, and should realize the importance of the extended family as a social and economic unit. Factors that could be discussed are the role of the traditional structures of authority, the ways of gaining access to land, and the major characteristics of the local farming system.

On-the-job training sessions are useful to the enumerator, since they broaden his knowledge and give him an opportunity to discuss his experiences with his colleagues. They can also be useful to the unit director, who can gain much knowledge if he is careful to keep the sessions informal. The enumerators should feel free to mention problems, to ask questions, and to voice their opinions on any aspect of their work; they should be allowed to criticize the survey methods and the questions used in the forms. They should also be given a chance to discuss the first draft of any new questionnaire; they are likely to make sound suggestions to improve its wording or layout.

One-day training sessions are easy to organize, require little equipment (a small blackboard and chalk is all that is needed), and can take place in the shade of a tree or in a shed in a village. If they are held regularly, they will become part of the enumerators' routine, so that there is less chance of them forgetting the exact date.

For example: The easiest way is to agree once and for all that the session for, say, the enumerators of Villages A, B, C, D and their supervisor will take place every first Thursday of the month in a central location.

14.3 Annual training sessions

Each year, before the start of the crop season, a training session should be organized for the entire staff of the unit. These annual training sessions enable the staff of the unit to get together to compare experiences and, very importantly, to benefit from the psychological boost of feeling that they belong to a group working for a common purpose.

For the enumerators especially, working in isolation in the field can be lonely and depressing, and filling in questionnaires can fairly quickly become a boring routine. Even if an enumerator enjoys good relationships with the farmers, he may have few contacts with people of his level of education. He may be worrying about employment opportunities after the project has ended and will feel that any extra training he receives will stand him in good stead.

For all these reasons, it is useful, both practically and psychologically, to prepare a training program that includes annual training sessions for the enumerators, supervisors, and data processors.

15 THE UNIT'S MATERIAL NEEDS

Farm monitoring programs that use many different questionnaires, special measuring instruments, or call for the enumerator to travel are likely to run into material difficulties. When the survey program and the work load of the enumerators are being decided, these material constraints should be taken into consideration. Designing good questionnaires and training the enumerators are intellectually rewarding tasks; organizing the mundane material support and regularly checking it are time-consuming and potentially boring tasks, yet they are essential to the success of the program.

The best possible questionnaire is useless if it does not reach the enumerator on time. An enumerator can find himself short of questionnaires because someone forgot to send him his regular supply, or miscounted the pages, or the person bringing them left them in the wrong village. If the enumerator does not have the necessary equipment when he needs it, he can miss vital items of information. All this can happen easily even in a well-organized unit. Each step of the survey program should therefore be considered in the light of: What can go wrong?

If at all possible, the budget should allow for goods to be purchased locally even if they are more expensive there than elsewhere. The problems of importing supplies (delays in shipping, customs clearance) will usually outweigh the price advantages and increase the risk of incomplete or incorrect data. The purchase of equipment should be planned well in advance; a delay in implementing a survey because of the lack of equipment can mean months of wasted effort.

For example: If the measuring chains and scales required for making yield estimates are not available when needed, the data gathered on the crop will have lost much of their value. The harvest will not wait!

15.1 Questionnaires

A farm monitoring program requires the preparation of many questionnaires. For the program to be respected, the office staff must ensure that the questionnaires are typed and duplicated, that they are distributed to the right enumerator at the right time, that the completed questionnaires are returned promptly to the office, and that they are properly processed.
When the questionnaires are being designed, care should be taken that they can be typed easily on the stencils and typewriters available at the office, and that the stencils are compatible with the available duplicating machines. Again the rule holds, what is available locally should be used whenever possible, even if it is more expensive. Ordering and shipping from abroad will always mean delays.

One of the office staff should be made responsible for the material organization of the questionnaires. This means he must ensure that the questionnaires are typed, duplicated, distributed, and collected in accordance with a precise program. This does not mean that he handles every page himself, only that he arranges for the work to be done on time and checks that it has indeed been done.

For example: He must know that he has to take the stencils of the year-round survey to the duplicating room by 15 March, 15 June, 15 September, and 15 December, and that he has to order 750 copies. It is his responsibility that the copies be ready by 1 April, 1 July, 1 October, and 1 January, that the pages be stapled per survey and put together for each enumerator (or that holes be punched if the pages are to be kept in a binder), and that they be ready for distribution by 15 April, 15 July, 15 October, and 15 January.

Packs of questionnaires are prepared for each enumerator. If distribution is likely to be difficult, the enumerators should be given a large supply at a time, say for three months. Each supply should be well packed; a large envelope is good if it is not too full; a plastic bag taped shut is better. Each pack should be clearly labelled with the name of the enumerator, his address, and the contents.

The packs will probably have to be hand-carried to the enumerators. It is best to have the supervisor take them when he goes to check the enumerator's work. Otherwise, the packs will have to be entrusted to other members of the project, in which case delays and risk of loss will be greater. In whichever way they are distributed, the person in charge of the questionnaires should keep a notebook indicating what was sent and who took it to each enumerator.

The problem of bringing the completed questionnaires back to the office is different: this *must* be done by the supervisor. And before doing so, it is imperative that he check the questionnaires with the enumerator for completeness and accuracy. This crucial step in ensuring the validity of the data was discussed in Chapter 7.

As soon as any questionnaires are brought to the office, this should be recorded in the notebook, so that it is always clear what has been sent and what has been received. The data processors should be instructed how to file the questionnaires received from the field.

The data will need to be accessible to many people. Filing the questionnaires is therefore more than just storing them out of the way. A system has to be developed for both the questionnaires and the recapitulative forms so that any form or item of data can be found quickly. Filing can be done by subject or by location, in binders or hanging files.

The system chosen for data filing must be known and clearly understood by all the people who handle the data; otherwise it will quickly become difficult to find anything. It is best to have one person in charge of checking all incoming questionnaires and filing them, as well as filing all recapitulative forms and keeping track of where each file is at all times.

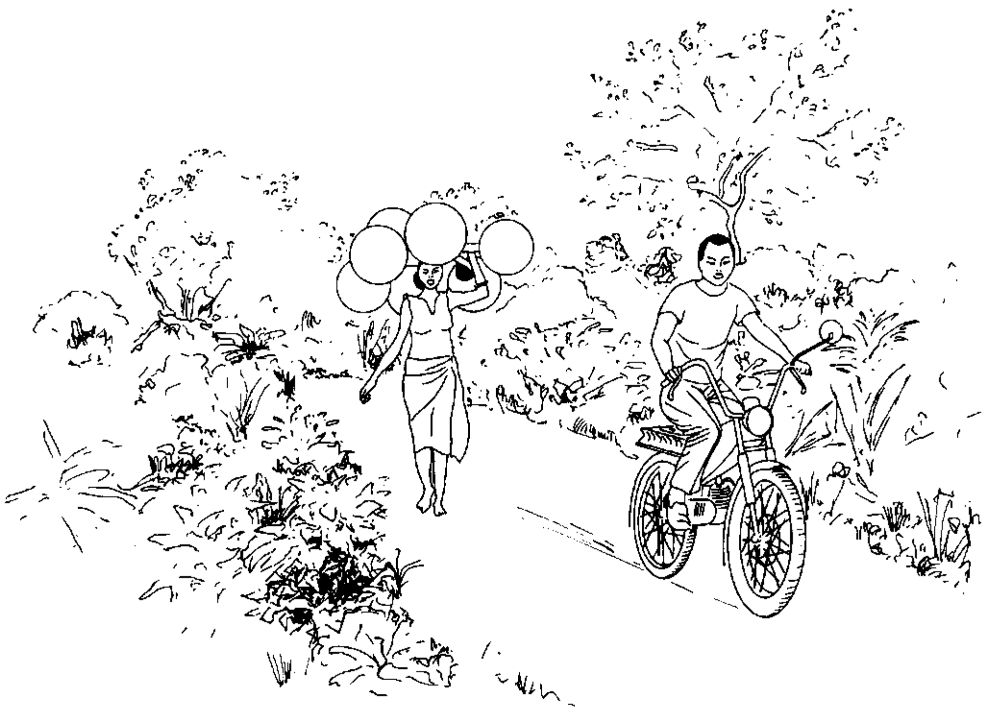
15.2 Material needs of the enumerators

15.2.1 Housing and transport

In a long-term survey, the enumerator can do a better job if he lives among the farmers whom he will be observing and interviewing. In small villages, however, it may not be possible to rent a house. A good solution in such a case is to have a house constructed for him and his family on the spot. The cost will be more than compensated for; it means that excessive travel is avoided, and that relationships between the enumerator and the farmers are likely to be better.
Alternatively, the enumerator could simply be given a housing allowance each month and told to find a place to stay. This has the advantage of being easy to organize but there is no control on where the enumerator lives. In practice, he will often go to the nearest large village.

If the survey is only of short duration (e.g. a reconnaisance survey) and the enumerator does not bring his family with him, he should be able to rent a room or a *case* with a farmer's family in the village. Renting a room with a farmer, however, can influence the survey. Should the landlord be included in the sample? Will staying with one family create problems in the inquiries with its neighbours? The enumerator should be very careful in his choice of lodging.

Wherever the enumerators live, they must be able to get to the farmers in the fields. Whether they can use bicycles or mopeds for their journeys depends upon the distances involved and the funds available. It should be very clear who owns the means of transport (the enumerator or the project), who pays for repairs and maintenance, and what is done with the bicycle or moped at the end of the project or when an enumerator leaves. A good system is for the project to advance the cash needed to buy the bicycle or moped and have the enumerator repay the loan through a salary withdrawal plan. Each enumerator should receive a monthly allowance to cover the depreciation and running costs, but he is responsible for extra repair costs. In this way, the bicycles or mopeds are likely to be better cared for, and the enumerators are more likely to be satisfied.



An enumerator must be able to get to farmers in the field

If the project buys bicycles or mopeds, it is safer to buy a brand popular in the country, so that mechanics and spare parts will be available even in rural areas.

15.2.2 Enumerators' equipment

The equipment issued to the enumerator has to be simple, sturdy, and as maintenance-free as possible. The following list is merely indicative and should be adapted to local conditions. Only cheap and simple devices are described here.

- *Chains and tapes*. Enumerators responsible for yield estimates need a chain or another measuring device. The chain or tape should be strong and easy to transport; tapes of poor quality are likely to break. Measuring wheels are easy to use but are fragile and the margin of error on irregular ground will be great.
- *Compass*. If fields have to be measured fairly precisely, the enumer-

ator will need a compass. This is a sensitive instrument and should only be provided if really needed. It has to be large and easy to read, with the degrees well spaced. The reading should remain correct even if the instrument is not held completely horizontal. The compass will need to be protected against dust and should be able to function well in high temperatures and high humidity. If these demands cannot be met or if the enumerator is not likely to be able to keep a compass operational, it is better for him to work without one.

- *Scales*. Independent of the method of yield estimates used (crop cutting or post-harvest estimate), the enumerators need a set of scales. It should be small and strong enough to be transported on the back of a bicycle or hand-carried. In the farmer's compound or in a field, hanging scales can be hooked onto a tree. A weighing capacity of fifty kilograms maximum will usually be sufficient. Essential features are easy readibility of the scale from various angles (as little parallax as possible) and a minimum of loose parts. Any loose part, such as a hook, will get lost easily so spare ones should be kept in stock, or it should be possible to use the scale without the part or to replace it with some wire.
- *Equipment for sample plot harvesting*. Each enumerator needs a couple of sacks in which to put the harvest and weigh it on the scale. If funds are available, there should be enough sacks so that each farmer can keep the sacks used for weighing his crops. A wooden or iron form (circle or square) can be very helpful when selecting small sample plots. All sample plots are marked off with pickets. These pickets can be cut by the farmer's family (for a fee). With bigger sample areas or tall crops it can be useful to connect the pickets with a rope.
- *Rain gauge*. If enumerators have to register the amount of rainfall, they need a rain gauge. Any type can be used as long as the scale is easy to read and is in known units of measurement (mm or inch). After the gauge has been installed, it should be protected so that animals cannot move or damage it.
- *Notebooks and binders*. Loose questionnaires get dirty, lost, or out of place. This is more so while they are with the enumerators than at the office. One solution is to staple all the inquiries of one type together in a sort of monthly or yearly booklet (only if staplers

that can handle thick piles of paper are available). Loose-leaf binders are another good solution. The enumerator can keep his questionnaires in several binders, colour-coded by household or by subject. The sheets should be strengthened around the holes to avoid being torn. All binders and punchers should be compatible. So that the enumerator need only carry one binder at a time, it might be best for him to keep together all the survey forms required for the households he visits on the same day.
The enumerators also need a notebook in which to write their observations during their trips to the fields.

- *Other equipment*. It is advisable to provide the enumerators with boxes or shelves in or on which to keep their forms and binders. This will help protect the documents from termites and damp.
  Each enumerator should be given a rainproof bag in which to carry his papers when he visits the farmers.
  If the interviews are conducted in the evening, the enumerators will need lamps. A battery-powered lamp on a stand is better than a flash light. The enumerator needs a regular supply of batteries, or money to buy them.

15.2.3 Enumerator's Manual

During the initial training session, each enumerator should receive a copy of the Enumerator's Manual, written especially for the surveys in the project. The Manual explains the objectives of the surveys, describes each questionnaire in detail, and outlines the daily and weekly activities of the enumerator. The supervisor should go over the Manual with the enumerators, showing them how to use it.
Each questionnaire should be described in detail, with the purpose of each question explained and instructions provided for filling in the answers. The Manual should review the problems likely to arise and advise on how to solve them.
The Manual should be updated and revised as the need arises (e.g. when additional surveys are included in the program or when an unexpected problem comes up frequently in a questionnaire). If loose-leaf binders are used, the Manual can always be modified and expanded.

15.3 Material needs of the data processors

15.3.1 Office equipment

The monitoring unit is likely to be the project's largest consumer of paper and the one most dependent upon office equipment. Getting enough questionnaires duplicated and sent to the enumerators on time is basic to the success of the surveys, so the unit should, if at all possible, have its own typewriter and duplicating machine. Enough stencils and paper should be stocked for a few months of questionnaires in case of shortages among suppliers.

Even in a small project, the quantity of completed questionnaires will be very large. The filing system therefore has to be organized from the very beginning, as otherwise much valuable information can be lost. Fire-proof metal cabinets that can be locked are best. Metal drawer-and-shelf systems are usually available but can be rather expensive. Alternatively, boxes can be built from plywood, with small shelves on which a pile of questionnaires fits easily (e.g. one box for one survey of one village for one crop season). When a survey does not use too many sheets of paper, they can be filed in large binders, grouped in whichever way is desired (per subject, per household).

A notebook with a list of the completed questionnaires and where they are filed will be of great help. To avoid the loss of documents, the completed questionnaires should be filed immediately they are received and the originals should never leave the office. If a photocopying machine is available, other units in the project can copy the data they need; otherwise any work they do on the data should be done in the unit's office.

15.3.2 Calculators

Each data processor needs a simple calculator to work out averages, percentages, areas, and yields. A calculator with a printer is most practical since it provides a written record of each step of the calculations. Battery-powered calculators might be the best choice. Calculators with rechargeable battery packs can be sensitive to voltage fluctuations so a voltage regulator might be advisable. The machines should be kept covered when not in use, and a large stock of printing paper kept on hand.

The medium chosen for the analysis of the data (hand or computer) depends in part on the quantity of usable data and on its quality. Whichever medium is chosen, the unit should have at least one programmable calculator for which standard programs for statistical analysis are available. Such calculators can cope with most of the calculations needed for preliminary reports without any help from outside the unit.

Small, cheap micro-computers are now available and can be used by people with little experience in computer work. These machines are capable of handling a great quantity of data and are likely to be sufficient for most monitoring units. They can be operated within the unit.

If the services of a computer centre are to be used, funds should be available for computer time, and either the unit director must have previous experience in computer work, or a programmer must be available to assist him.

16 BACKGROUND INFORMATION REQUIRED BY THE UNIT

Before the unit director can prepare the survey program and choose appropriate methods of data gathering and analyses, he has to learn about the natural and human context within which the project is taking place. He also needs a detailed description of the project program and of the results it expects to achieve, as well as basic information on the project population.

16.1 The context of the project

To understand the context of the project, the director needs data on the geography of the area, its administrative structure, social organization, economic activities and resources.
Much of this information can be found in national statistics, in regional administrative records, or in reports of earlier projects or studies done in the region. Sometimes, valuable information can be obtained from interviews with knowledgeable people or through group interviews.

16.1.1 National statistics

Statistics at the national (and sometimes regional) level are available from government publications and from those of such international agencies as the Food and Agriculture Organization (FAO) and the World Bank. In spite of a constant effort to unify standards and methods,

discrepancies are usually found in such documents. The quality of statistics put out by individual governments, in particular, can vary greatly, so the director should try to find out how the data were gathered and processed, and whether they can be considered valid for the unit or not. Valid data can be used by the monitoring unit even if they are grouped differently from the way required for project reporting. It is worth accepting a margin of error due to the regrouping of data if the data are valid to start with.
Comparisons of data from different sources should be done with great care to ensure that only comparable data are compared. Comparisons of prices from countries with different currencies, for instance, are meaningless unless adjustments are made to express all prices in the same currency. Also misleading is to compare changes in prices over time, even in the same country, unless inflation is taken into account.

Some institutions that can be valuable sources of information are listed below. Most of them have offices in countries of semi-arid West Africa. This list is not meant to be exhaustive; it is given merely as a guide.

- Comité Inter Etats de Lutte contre la Secheresse au Sahel (CILSS);
- European Development Fund (EDF);
- Food and Agriculture Organization of the United Nations (FAO);
- Institut de Recherche Agronomique Tropicale et des Cultures Vivrières (IRAT);
- Institutions dealing with specific crops;
- International Bank for Reconstruction and Development (IBRD);
- International Centre for Agricultural Research in Dry Areas (ICARDA);
- International Crops Research Institute for the Semi-Arid Tropics (ICRISAT);
- International Food Policy Research Institute (IFPRI);
- International Institute for Tropical Agriculture (IITA);
- International Maize and Wheat Improvement Centre (CIMMYT);
- Marketing and price control institutions;
- Ministries and dependent institutions such as the Extension Service and Rural Credit Units;
- National research centres;
- Office de la Recherche Scientifique et Technique Outre-Mer (ORSTOM);

- Organization for Economic Cooperation and Development (OECD);
- United Nations Development Project (UNDP);
- United Nations Educational, Scientific, and Cultural Organization (UNESCO);
- United Nations International Children's Emergency Fund (UNICEF);
- Universities and colleges of agriculture;
- World Bank;
- World Health Organization (WHO).

Much general agricultural information on Africa is available in the French *Mémento de l'Agronome* (available from the *Ministère de la Coopération, 20 Rue Monsieur, 75700 Paris, France)*. Three publications, in particular, present basic national statistics on any country. They are:

- The yearly World Development Report of the World Bank;
- The production yearbooks of FAO;
- The Agricultural Census of FAO.

As an example of data available, part of a table from the 1980 World Development Report of the World Bank is reproduced in Table 16.1.

16.1.2 Regional administrative records

Information available in the regional administrative records is likely to include demographic data, lists of villages and their boundaries, infrastructure, the locations of schools and dispensaries, official market places, and details of extension services and credit units.

Before using data from such sources, the director should check why the data were gathered in the first place and whether this may make them incompatible with what the unit intends to use them for. This can easily happen.

For example: Suppose that the monitoring unit needs a list of the heads of households in an area. Lists available from the regional administration are often prepared for tax purposes. Such lists might include people born in the area who are not physically present in a village, whereas recent settlers may still be listed in their village

Table 16.1. Example of a table from a World Bank publication

	Population (millions) Mid-1978[a]	Area (thousands of square kilometers)	GNP per capita: Dollars 1978[a]	GNP per capita: Average annual growth (percent) 1960–78[b]	Average annual rate of inflation (percent) 1960–70[c]	Average annual rate of inflation (percent) 1970–78[d]	Adult literacy rate (percent) 1975[e]	Life expectancy at birth (years) 1978	Average index of food production per capita (1969–71 = 100) 1976–78
Low-income countries	**1,293.9** *t*	**26,313** *t*	**200** *w*	**1.6** *w*	**3.0** *m*	**10.6** *m*	**38** *w*	**50** *w*	**97** *w*
1 Kampuchea, Dem.	*8.4*	181	..	..	3.8	..	..	..	57
2 Bangladesh	84.7	144	90	−0.4	3.7	17.9	26	47	90
3 Lao PDR	3.3	237	90	..	..	..	..	42	96
4 Bhutan	1.2	47	100	−0.3	..	..	..	41	100
5 Ethiopia	31.0	1,222	120	1.5	2.1	4.0	*10*	39	84
6 Mali	6.3	1,240	120	1.0	5.0	7.8	10	42	90
7 Nepal	13.6	141	120	0.8	7.7	9.1	19	43	92
8 Somalia	3.7	638	130	−0.5	4.5	10.7	60	43	87
9 Burundi	4.5	28	140	2.2	2.8	10.1	25	45	107
10 Chad	4.3	1,284	140	−1.0	4.6	7.4	15	43	89
11 Mozambique	9.9	783	140	0.4	2.8	10.9	..	46	81
12 Burma	32.2	677	150	1.0	2.7	13.7	67	53	96
13 Upper Volta	5.6	274	160	1.3	1.3	9.6	*5*	42	95
14 Viet Nam	51.7	330	170	..	..	..	87	62	102
15 India	643.9	3,288	180	1.4	7.1	8.2	36	51	100
16 Malawi	5.7	118	180	2.9	2.4	9.1	*25*	46	99
17 Rwanda	4.5	26	180	1.4	13.1	14.7	*23*	46	103
18 Sri Lanka	14.3	66	190	2.0	1.8	11.8	*78*	69	114
19 Guinea	5.1	246	210	0.6	1.7	*6.4*	..	43	86
20 Sierra Leone	3.3	72	210	0.5	2.9	10.8	15	46	93
21 Zaire	26.8	2,345	210	1.1	29.9	26.2	15	46	94
22 Niger	5.0	1,267	220	−1.4	2.1	10.7	8	42	87
23 Benin	3.3	113	230	0.4	1.9	7.4	*11*	46	92
24 Pakistan	77.3	804	230	2.8	3.3	14.6	21	52	101
25 Tanzania	16.9	945	230	2.7	1.8	*12.3*	*66*	51	93
26 Afghanistan	14.6	647	240	0.4	11.9	4.4	12	42	100
27 Central African Rep.	1.9	623	250	0.7	4.1	9.0	..	46	102
28 Madagascar	8.3	587	250	−0.3	3.2	9.6	*50*	46	95
29 Haiti	4.8	28	260	0.2	4.1	12.2	23	51	91
30 Mauritania	1.5	1,031	270	3.6	1.6	10.4	*17*	42	71
31 Lesotho	1.3	30	280	5.9	2.5	11.2	*55*	50	90
32 Uganda	12.4	236	280	0.7	3.0	27.3	..	53	90
33 Angola	6.7	1,247	300	1.2	3.3	22.0	..	41	88
34 Sudan	17.4	2,506	320	0.1	3.7	7.4	*20*	46	108
35 Togo	2.4	56	320	5.0	1.7	7.4	18	46	80
36 Kenya	14.7	583	330	2.2	1.6	12.0	*40*	53	91
37 Senegal	5.4	196	340	−0.4	1.7	8.0	*10*	42	96
38 Indonesia	136.0	2,027	360	4.1	..	20.0	62	47	100
Middle-income countries	**872.8** *t*	**32,998** *t*	**1,250** *w*	**3.7** *w*	**3.1** *m*	**13.1** *m*	**71** *w*	**61** *w*	**106** *w*
39 Egypt	39.9	1,001	390	3.3	2.7	7.0	*44*	54	93
40 Ghana	11.0	239	390	−0.5	7.6	35.9	30	48	79
41 Yemen, PDR	1.8	333	420	..	..	..	*27*	44	108
42 Cameroon	8.1	475	460	2.9	3.7	*9.8*	..	46	112
43 Liberia	1.7	111	460	2.0	1.9	9.7	*30*	48	96
44 Honduras	3.4	112	480	1.1	3.0	8.0	57	57	84
45 Zambia	5.3	753	480	1.2	7.6	5.7	*39*	48	109
46 Zimbabwe	6.9	391	480	1.2	1.3	7.6	..	54	102
47 Thailand	44.5	514	490	4.6	1.9	9.1	84	61	122
48 Bolivia	5.3	1,099	510	2.2	3.5	22.7	*63*	52	111
49 Philippines	45.6	300	510	2.6	5.8	13.4	*87*	60	115
50 Yemen Arab Rep.	5.6	195	520	..	..	..	*13*	39	98
51 Congo, People's Rep.	1.5	342	540	1.0	5.4	10.6	50	46	82
52 Nigeria	80.6	924	560	3.6	2.6	18.2	..	48	89
53 Papua New Guinea	2.9	462	560	3.6	3.6	8.8	32	50	106
54 El Salvador	4.3	21	660	1.8	0.5	10.3	*62*	63	111
55 Morocco	18.9	447	670	2.5	2.0	7.1	*28*	55	80
56 Peru	16.8	1,285	740	2.0	9.9	22.2	72	56	90
57 Ivory Coast	7.8	322	840	2.5	2.8	13.9	20	46	104
58 Nicaragua	2.5	130	840	2.3	1.9	*11.0*	57	55	102
59 Colombia	25.6	1,139	850	3.0	11.9	21.7	*81*	62	114
60 Paraguay	2.9	407	850	2.6	3.0	12.3	81	63	103
61 Ecuador	7.8	284	880	4.3	..	14.8	74	60	103
62 Dominican Rep.	5.1	49	910	3.5	2.1	8.6	67	60	93
63 Guatemala	6.6	109	910	2.9	0.1	10.8	47	57	108
64 Syrian Arab Rep.	8.1	185	930	3.8	1.9	12.7	53	57	150
65 Tunisia	6.0	164	950	4.8	*3.7*	7.1	55	57	128
66 Jordan	3.0	98	1,050	..	..	..	70	56	77

Source: Table 1 from the World Development Report 1980. New York, Oxford University Press for the World Bank, Washington D.C., August 1981

Source: Table 1 from the World Development Report 1980. New York, Oxford University Press for the World Bank, Washington D.C., August 1981

of origin. Lists from extension services or credit units may also be biased, since they will probably only include the names of those participating in their programs.

These warnings do not mean that such records can never be used to obtain background information, only that the director should be aware of the bias inherent in them (because of the purpose for which they were compiled) before deciding whether and how he can use them.

16.1.3 Reports from other projects

Much can be learned from the reports of other projects, whether projects implemented for the same general purpose as the project that will be monitored, or simply projects implemented in the same region. Project reports are not always easy to obtain but it is worth spending time trying; the types of problems encountered in previous projects are likely to arise again.
Projects in the area can be identified by inquiring at the ministries concerned, or from funding agencies (inquire at embassies and at the headquarters of international institutions). Similar agricultural development in other areas may be more difficult to identify. If a library is available, it is worth looking systematically through its catalogues or card systems, under subject headings of interest but also under geographical locations similar to that of the project.
Reading the reports of other projects can be helpful in several ways:

- They provide useful information on the way the local people farm and on the way they react to new technology;
- They enable a study of what methods were used for data gathering and analyses and how the results are presented in the report. On this last point, bad reports are as educational as good ones!

16.1.4 Interviews with knowledgeable people

Much useful background information is not written down but is known to people living and working in the area. Local administrators, traditional chiefs, extension agents, or researchers can all be useful sources of information, although what they say should be checked against other available information.

Interviewing knowledgeable people

The unit director should plan to visit the project area and to talk at length with local people. At first, informal conversations are best, just asking some general questions to get the person to talk. After a few such interviews, it may be found that some people contradict one another. This identifies subjects that should be investigated in more depth.
Some researchers have had good results with group interviews, sitting down with a few farmers or extension agents and informally discussing the local situation. This is best done with people of about the same social status, who will feel free to speak their minds in front of the others. A group-interview with older and younger farmers to talk about changes in agricultural practices can be useful, provided that the young people are allowed to speak up in front of their elders.

16.2 Data on the project

16.2.1 The project program

If the monitoring unit is to find out how the farmers are reacting to the improvements being introduced by the project and what results they are achieving, it must have a detailed description of the project program and of the expected results. This is not always readily available from project documents, so the director may need to interview project specialists to prepare the description. The items that need to be described depend upon the project, but some suggestions are given below:

- For whom is the project being implemented? Many projects are geographically defined, so they may be directed to all the households in the area. Other projects are restricted to farms that grow a certain crop or share some other characteristic;
- How is the project to be implemented? This description should cover the methods that will be followed and the means available to do so. It should supply answers to such questions as:
  - Will the extension agents systematically visit all the farms in their area, or only a selection of farms, or only farms to which they are invited?
  - If inputs and equipment are being promoted, where will these be available? At what prices?
  - Will there be a credit service? If so, to which farmers will it be available?
  - What measures will be taken to market the crops?
- What is the project trying to achieve? This should be answered in concrete terms. To increase food production is a commendable aim but it is not precise enough. An ideal statement would be: to increase average sorghum yields from 500 kg/ha to 800 kg/ha.

Sometimes a project's original program changes during implementation and the program really carried out is not put in writing. The unit director must see to it that he is kept informed of such changes by maintaining regular contacts with the specialists concerned. He should then write a summary of the changed program and discuss it with the project management so all agree that it is correct.

16.2.2 Data available from other sections of the project

Data on project implementation will be gathered routinely by other sections of the project. Items likely to be included are the type, price, and quantity of fertilizer and other inputs available to the farmers, the amount of credit taken up by each farmer and for what purpose, and the completion of infrastructure such as wells and roads.

The monitoring director should maintain regular contacts with the heads of other sections, inquiring about the data that they can provide him with and arranging for the data to be passed on to his unit. This should be organized right away, as otherwise chances are that no one will take the initiative of sending the data. The surest way is to get an authorization from each section head, allowing one of the unit's data processors to visit his section regularly to copy any relevant data.

16.3 List of a project population

A population list required to select samples is not likely to be available, but it may be partly available. When the population to be studied consists of farms, farmers, or households, it is likely that some sort of official list will be available in the regional administrative records (16.1.2). Although such lists may not be complete, it is usually easier to correct them than to start over again from scratch. The local administrators or the traditional chief will be able to say which people on the list are not actually present in the village, and which households have recently settled in the area. A simple walk through a small village, with the director inquiring from the local leaders which household lives in which homestead, can suffice to complete a list.

16.4 Case studies

Before the unit director can finally decide on the methods of data gathering and select the samples that will be surveyed, he needs basic data on the project population. These can be obtained through case studies conducted on a few families, which will give him a good idea of the local way of life.

A case study means interviewing the members of a household in great detail to obtain a complete picture of their life. All members of the household should be interviewed, not just the men. Case studies can only be done on households whose members are willing to spend a lot of time on interviews and do not mind talking about their personal affairs. The people interviewed will thus not be truly representative of the local people, but their answers will throw some light on the subjects that should be included in the surveys and will indicate what methods would be practical for data gathering.

For example: After talking with a few farmers, the director will know whether the farmers in the area usually bring their cereal harvest to their homestead each day or whether they store it on the field until the entire field has been harvested. This information will allow him to choose the best method of estimating field production.

16.5 Updating the information

The background information that has been gathered to prepare the monitoring program will need to be kept up to date. The unit director should keep informed of any changes in services available locally through private and official organizations and of any changes in official prices and marketing channels.

For example: A mill may be opened in a village, or the cereal board may obtain more trucks and thus be able to buy the crops in more villages than before.

It is important to find out right from the beginning what data will be available, when, and in what format, so that the unit can avoid duplication of effort while still obtaining all the information it needs.

17 SELECTION OF A SAMPLE

Data will usually not be gathered on the entire population, but on a sample which is selected to represent that population. This means that what is found out about the sample also applies to the entire population. This chapter will discuss only those sampling methods which are used to select a random sample.

To be representative, a sample has to fulfil two requirements:

- It has to be large enough;
- It has to be selected in accordance with certain rules.

17.1 Sample size

It is not possible to give precise rules on what sample size is the best. It does not depend on the size of the population, nor does it have to contain a minimum percentage of that population. Factors to be kept in mind when deciding the size of the sample are:

- The sample size should never be so large that it exceeds the work capacity of the enumerators. A safety margin should even be included, because delays in data gathering are unavoidable;
- If great differences between the elements of the population are expected, the sample has to be large. If those differences are expected to be small, the sample can be small. An extreme case would be a population with all elements identical, in which case a sample of one element only would suffice;

- The precision required by the users also plays a role; higher precision requires larger samples;
- Missing or wrong responses by the farmer and incorrect observations by the enumerator have to be rejected from the analysis. The sample that then remains to be analyzed will be reduced by an unknown percentage. In general, more complicated surveys have a higher percentage of data that must be rejected;
- As a rule of thumb, samples of at least ten to fifteen elements per subgroup should be available for analysis. It is emphasized that the percentage of rejected data should not be underestimated. Especially in the beginning of monitoring, the samples should therefore be as large as can be handled.

17.2 Random sample

A sample can be chosen by selecting some of the elements from the population at random, which means that all the elements of the population have the same chance of being selected.
To select a random sample, the population from which the sample will be selected first needs to be identified and a list of the elements of the population (farmers, fields, or whatever) has to be drawn up. Perhaps such a list already exists. Before it can be used, however, it should be checked to make sure that it is complete. As discussed in Chapter 16, official lists may be incomplete or biased because of the purpose for which they were drawn up.
If no acceptable list is available, it may be possible to combine lists. If not, a small additional survey will have to be made to identify the elements of the population from which the sample is to be drawn. Conducting a full census on every household in the area would be far too time-consuming and expensive, but a simple census, limited only to that information required to select the sample, could be done.

Once a list of the elements of the population is ready, each one is given a number. The sample can then be selected in any of three ways, all resulting in a random sample.

17.2.1 Random sample on a small population

If the population is small, the numbers of the population elements can be written on small pieces of paper, one for each element. These papers are put in a hat or a box and mixed together. Without looking at the papers, somebody takes out as many as are needed for the sample. This is the same system as is used for some lotteries. It has the advantage that it can be done in front of the farmers being chosen for the sample, to show them that the choice is really at random, without preference for certain people.

A random sample is being selected

17.2.2 Random sample on a large population

If the population is large, the above approach is time-consuming. To avoid the work of copying the numbers of the population elements on small pieces of paper, a table of random numbers can be used. Table 17.1

is such a table. It groups the digits 0-9 in an arbitrary order, with each digit represented about the same number of times. How to use this table to select a random sample will be explained below.

Using a table of random numbers to select a random sample

Suppose the population consists of 576 households and a sample of 25 households is needed for a survey. All the households are given a three-digit number (001 to 576). To select the sample with the table of random numbers (Table 17.1), decide on any point in the table as a starting point (it does not matter where), then read the digits in a continuous line, horizontally, vertically, or diagonally, ignoring the blanks between columns of digits.

As each element of the population has received a three-digit number, units of three digits will be read from the table. In this example, the starting point was chosen at the beginning of Line 10 and the table was read along the horizontal line. The first unit of three digits is 880, which does not exist in the population. Nor does 828, the next three-digit unit, so both numbers are disregarded. The next number, 441, exists in the population and is chosen as the first element of the sample. The next three-digit number, 902, is examined and rejected because it is too big, and so on. At the end of the line, the digits are read on the next line without interruption (443).

The numbers chosen are listed below; those not existing in the population and therefore disregarded are placed between parentheses. If a number is selected twice, it is disregarded the second time and is also placed in parenthesis.

(880), (828), 441, (902), 310, 315, (699), 545, (608), 487, (612), 181, (924), 314, (704), (836), 443, (765), 084, 533, (940), (668), 242, 568, (634), (871), (659), 145, 467, (924), (892), (775), 124, (670), (941), 320, (761), 127, 030, 482, 509, (871), 225, 016, 421, 262, (774), (951), (583), (739), 370.

It is sensible to select a few elements more than are needed so that if any have to be rejected for one reason or another, their replacements have already been chosen. The random numbers system is simple, and guarantees the absence of biases in the sample.

Table 17.1. Random numbers

	1	2	3	4	5	6	7	8	9	10
1	35994	98850	60645	47230	11263	99741	03582	30302	19643	77763
2	35344	66337	23823	72763	22369	99807	70900	55238	83395	15261
3	13823	62839	94098	55189	95052	37627	18776	00487	56759	50748
4	98609	33844	18419	46382	21125	36269	44139	82930	88609	68568
5	85899	92331	16951	47097	50041	80599	65130	99875	26349	81152
6	29189	54644	98913	51947	75933	15507	35643	89179	42506	70953
7	05702	91196	67141	07335	94480	21070	74326	54236	75337	25949
8	91229	76151	43632	08102	13994	15529	56338	61738	64360	79962
9	74455	84531	20710	21288	17400	26916	23756	14270	25379	13970
10	88082	84419	02310	31569	95456	08487	61218	19243	14704	83644
11	37650	84533	94066	82425	68634	87165	91454	67924	89277	51246
12	70941	32076	11270	30482	50987	12250	16421	26277	49515	83739
13	37089	92630	79234	94344	58856	19405	17165	90585	58610	62884
14	55793	45521	79262	41407	89092	43135	94170	97859	11158	70937
15	12016	83870	37223	16277	58520	63383	59563	03050	31501	02558
16	74210	26306	76189	94150	78056	45999	10514	53639	10769	14270
17	82612	66581	88352	94158	81060	71195	20366	83292	95216	49271
18	24686	48363	80991	88500	97464	04812	96211	99007	45035	32276
19	55425	41820	76775	47193	81795	05076	46937	60088	87967	53353
20	91748	41776	06702	59071	40726	83102	56119	63047	33845	29542

From: *Inleiding tot de Statistiek*. H. Rijken van Olst. Van Gorcum & Comp. N.V., Assen, 1966.

17.2.3 Systematic random sample

A systematic random sample can be chosen directly from the population list. The procedure is as follows:

- Decide upon the size of the sample;
- Decide what interval to leave between elements chosen for the sample; this interval should be about equal to the size of the population divided by the sample size so that the whole list is considered for the sample;
- Take the list of the population elements and choose an arbitrary number to indicate which element at the beginning of the list will be the first element of the sample. This number has to be smaller than the size of the population divided by the sample size;

- Go down the list, choosing elements at that interval.

Alternatively, one can start in the middle of the list and choose elements on either side of this middle.

Example of systematic random sampling

There are 245 households in the villages and a sample of 25 households will be studied. To find the interval between the elements chosen, we divide the population size by the sample size; this gives us 9.8. The number 10 would be an easy choice, but that would give us only 24 elements in the sample when we arrive at the end of the list. Therefore we choose 9 to be the interval size. We then have to choose a number smaller than 9 to indicate where to start sampling in the list. Let us take 8. The elements selected for the sample are then Households 8, 17, 26, ... 215, 224.

Alternatively, if it is decided to start in the middle of the list, the interval can be 10. Household 123 in the middle of the list is chosen to start with, and then every tenth household on either side of 123, twelve on each side. So Households 133, 143, 153, and so on are taken down the list, and 113, 103, 93, and so on are taken up the list.

A problem with this method is that it requires a careful check for any pre-existing periodicity in the list. Periodicity means that elements with a common characteristic are recorded at regular intervals on the list. If that interval coincides with the interval chosen for the sample selection, the sample will consist only of elements with that characteristic, or no elements at all with that characteristic. Any built-in periodicity in the list is only relevant if it concerns a topic being considered in the survey.

For example: Suppose that orange traders and mango traders are placed alternately at the market place, side by side. If it is decided to take a systematic sample of traders with an interval of 4 and one starts with a mango trader, only mango traders will be chosen; or if one starts with an orange trader, only orange traders will be chosen.

The problem of periodicity can sometimes be solved by selecting a sample larger than that required by the systematic sampling method and then drawing the final sample from the large sample. Lists with built-in periodicity, however, are rare, so the systematic sample method, which is simple and fast, can usually be applied.

17.3 Two-stage sampling

In many projects, samples cannot be selected completely at random; there may not be enough enumerators to monitor a sample spread out over a large area, or not all the villages can be reached in the rainy season. To avoid these problems, a two-stage sampling procedure can be used.

The first stage is to select locations where surveys will be conducted. If enough enumerators are available, as many locations can be chosen as there are regions to be represented. Those locations can be villages selected at random from a list of all the villages, although it is better to select them so that they are more or less representative for the topics being considered by the surveys. These villages must be accessible all year round, although this very fact may mean that they are different from the other villages in the conditions under which the farmers work (supply of inputs, markets etc.).
The second stage is to choose a random sample in each of the villages chosen, using one of the methods described above.

17.4 Stratified sample

The greater the differences between the elements of a population, the larger the sample has to be to be truly representative. If there are obvious differences in the population, it can be divided (i.e. stratified) into sub-populations of like groups (i.e. strata) and a sample can then be taken from each stratum. The sum of the elements in the stratified samples can be less than the number of elements in the unstratified sample, while representing the population equally well.

Even if nothing is known about the population, it is often safe to assume that differences exist between large and small farms, or between modern and traditional farms etc. A simple census will probably confirm this assumption. With that information, the population can be divided into strata, after which a sample is taken from each. All strata are then represented in the sample according to a predetermined quantity. The main advantage of pre-stratifying the sample is that the different strata can be analyzed separately as well as together. If the sample is selected at random and stratified afterwards, some of the strata may be too small to analyze.

The characteristics upon which the population will be stratified depend, for a great part, on the analyses that are desired. These should be discussed with the users. Some of the characteristics often used to stratify a population of farmers are:

- Number of workers in the household; classification of numbers of workers in households depends on local conditions;
- Amount of land available to each household; this is feasible only in areas where such information can be obtained easily (e.g. where there is a system of land registration);
- Cash crop cultivation; sometimes a distinction has to be made between strictly subsistence farmers and farmers who cultivate some cash crops (e.g. in a project promoting a cash crop). Usually cash crops like cotton, peanuts, coffee and so on are marketed through an official board; if so, the board can provide a record of farmers from whom it bought some of those crops in previous years;
- Type of technology used; projects with an agricultural extension component often use a stratification based on a level of technology in which the farmer has some previous experience. This can include the use of improved varieties, fertilizer, animal traction, or any technique being promoted by the project. Such information may have to be obtained directly from each farmer.

Example of stratification

In a project which introduced animal traction, a request has been made for farm income data. The users need this information to compare the incomes of modern farmers (those using animal traction) with the incomes of traditional farmers (those not using animal traction). From this, they will determine the repayment capacity of the modern farmers for their equipment. In consultation with the users, it was decided that the ownership of a plough was a good indicator of a modern farmer. The population of farmers will therefore be stratified according to the ownership of a plough. First each farmer is interviewed to find out whether he owns a plough, and this information is checked against the list of the rural credit unit. A farmer can own a plough without being on that list, if he paid cash for it or inherited it. But if a farmer is on the credit list and says he does not own a plough, the information should be verified; he may well have sold his plough, and so should be in the no-plough stratum.

The sizes of the sub-samples taken from each stratum of the population do not have to be proportional to the size of the stratum they represent. However, the sub-samples should not be smaller than a minimum size, in line with the requirements explained for regular samples (17.1).

For example: Suppose that 20 per cent of the farmers in the above example own a plough. The enumerators can survey a total of sixty farmers. Assuming that the variations in farm income in both strata are the same, a sample of thirty farmers can be chosen from each stratum. If the two samples are to be combined to generalize for the total population, the results of the no-plough sample should be given a weight of 0.8 while the results of the sample with plough are given a weight of 0.2. This brings each stratum back to its correct proportion in the population.

17.5 Combining samples

If interactions between data are to be found, all the data have to be gathered on the same sample.

For example: If the influence of fertilizer on maize yield is to be found, the quantity of fertilizer used and the quantity of maize harvested should be surveyed on the same sample of fields.

Surveys covering different categories of data can be done on different samples. An advantage of using different samples is that the farmers are interviewed less often and with fewer questions, which is likely to produce better quality data. An extra advantage is that with fewer data being gathered from each farmer, there will be fewer farmers rejected from the analysis because some of their data are missing.

A sample can be chosen for each specific survey, although sometimes, when an additional survey is requested, all that is needed is to add some extra questions to a sample already being surveyed. This can save much time and money, but the sample already being surveyed should be examined to make sure it is acceptable for the additional survey.

For example: In an area where cattle raising is rarely practised, a simple random sample is being surveyed on the farmers' agricultural income. The monitoring unit is then asked to estimate the importance of income from cattle raising. Here, a few additional questions put to the farmers in the existing survey can provide the requested information. If, however, the request is to measure the income from cattle of farmers specialized in livestock, the old sample will not be adequate and a new sample has to be selected from among farmers specialized in livestock.

18 MARKETING

When one wants to compare the results achieved with different crops, it is meaningless to compare the quantities harvested. The same thing applies when comparing the results achieved on different farms. The best way to compare such things is to compare their value in money. To do this, a fair price must be chosen for each item. This is a most important choice, because the prices used will greatly influence the conclusions drawn by the monitoring unit. Using the same harvest data but different prices, one can easily draw opposite conclusions on the profitability of a crop.
To obtain data on prices, the monitoring and evaluation unit will need data on three aspects of the marketing channels open to the farmers:
- To or from whom can agricultural produce be sold or purchased?
- In what quantities?
- At what prices?

18.1 Marketing channels

18.1.1 The sale of produce

The farmer is likely to have various marketing channels open to him for the sale of his produce:
- Official government marketing boards;
- Private traders at the local market or at the farmer's home (large quantities);
- Other farmers at the local market (small quantities) or in the

farmer's own village (incidental);
- Marketing cooperatives.

Marketing boards buy only the main cash and food crops. Sometimes there is a separate board for each crop. These boards usually pay a fixed price all year round. The unit director should inquire how the organization functions and whether it operates throughout the country or only in large centres.

Private traders come to the local markets or to the farmers' homes at harvest time to buy produce. Except for the official marketing boards, this is the only way a farmer can sell produce in large quantities.

The farmer is able to sell small quantities at any time at the local market. Often the price he obtains there is higher than he would receive from private traders. But prices fluctuate with the season, and the seller usually has to pay a market tax. If the market prices are not officially controlled, the prices obtained will depend entirely on the supply and demand at that moment. If there are many buyers and little produce, prices will go up. The farmers may also sell some of their surplus produce to other farmers in their village, but this will only be on a small scale.

Marketing cooperatives can be very diverse in size and structure. If a cooperative exists in the project area, the unit director should inquire about its conditions for membership, the prices it pays, and the facilities available for storage and transportation.

Another form of a cooperative is a cereal bank, which is formed by a group of farmers who build a common storage facility. Part of the stock can be loaned to needy households, who will return a slightly larger amount after their next harvest. Some of the stock can also be sold long after harvest when prices are higher.

18.1.2 The purchase of inputs

Like the marketing channels for the sale of his produce, the farmer will have various channels through which to purchase his inputs:

- Agricultural credit unit or extension service;
- Private traders, usually at the local market;

- Individuals in the village (rare);
- Makers of agricultural tools (e.g. a blacksmith).

If an agricultural credit unit or an extension service exists, most necessary inputs can probably be purchased there (on credit). This is often the only way that expensive agricultural equipment like ploughs can be purchased. If the project is promoting animal traction, draught animals are likely to be available through these channels, or can be purchased direct from the nomads who raise them.
At local markets in West Africa, inputs such as fertilizer and insecticides are for sale only in small quantities, and these are mostly leftovers from farmers, or part of what farmers bought on credit and are now selling to obtain some cash. Other inputs likely to be available at the local market are hand tools.
In the village, farmers may sometimes be able to purchase or exchange small quantities of fertilizer or insecticide.
If there is a blacksmith in the village, they may be able to buy hand tools direct from him.

18.1.3 Transportation to market places

The means of transport available to carry produce to and from the market plays an important role in the use of marketing channels. If the farmers have easy access to some means of transport, they can go to markets further away if the price there is better. On the other hand, markets that cannot be reached by trucks will not be attractive to large-scale buyers, so trade there will be restricted to small quantities.

When the marketing channels available to the farmers are being monitored, not only the markets they visit should be studied, but also the means and costs of transport to the markets. To be able to compare the prices at the local market with those at nearby regional centres, one must know the transport costs to both places.

Transport to market

18.1.4 Marketing habits of the farmers

Trying to get complete data on the marketing of agricultural produce is an impossible exercise. It is a sensitive subject and many small transactions are likely to be forgotten so the margin of error would be high anyway. But what matters for the project is not to know whether a farmer sold 300 or 350 kg of cereals last year; the important thing to know is what happens when a farmer has a few hundred kilos of surplus? Does he store it all or does he sell most of it? Can he find a buyer for it if he wants to sell, and if so, when, where, and at what price?

A single-visit reconnaissance survey can be conducted on a sample of households, during which both men and women are interviewed (women often do much of the marketing at the local market) to find out which channels they use when they want to sell some produce.

When the director is compiling the questionnaires for this survey, he should avoid broad vague questions such as: 'Can you find a buyer for your cereals?'. Instead he should ask:

- Since the last harvest, have you sold any white sorghum?
- If so, to whom?

- At what period of the year?
- What was the price then?
- Where did the sale take place?
- How did you carry the produce to the selling place?
- How did the buyer take it away?

Each time a transaction is identified, the category of buyer should also be identified (marketing board, trader, farmer, etc.) and how the produce was transported to the place of transaction. In this way, the relative importance of the diverse marketing channels can be ascertained and plans can be made to improve them, if necessary.
As prices and marketing possibilities vary throughout the year with supply and demand, the timing of the transaction should also be noted.

18.2 Choice of prices

Knowing the marketing channels and the marketing habits of the farmers is helpful in choosing the prices to use when calculating the value of agricultural production. Those prices should, as far as possible, be those that the farmer is likely to receive, since it is those that he uses when planning his cropping pattern and agricultural practices.

18.2.1 Prices of inputs

The official prices of inputs (with or without credit costs) are a good approximation of their cost to the farmer, because most modern inputs are only available via credit units or extension services.
Small agricultural tools can be purchased at the local market as well, so these prices should also be obtained. They are generally fairly stable throughout the year.
The price of draught animals purchased from pastoral nomads varies not only with the age and weight of each animal but also with the season, so only rough estimates can be made for the price of an 'average' animal of a certain age.

18.2.2 Prices for the sale of produce

Government marketing boards purchase agricultural produce at the official prices, which are generally low. Usually the boards have a monopoly on the purchase of agricultural produce and are the only channels through which the farmers can sell their crops in large quantitities. If the farmers have a choice, they prefer to sell to private traders who give higher prices. The farmers could obtain even higher prices by selling to individuals at the local market, but this is possible only for small quantities at a time.
Farmers who are short of food crops may have to buy some at the local market at the current price; this is likely to be quite high because supplies will be low at the time most farmers start buying. To estimate the value of a crop, all these different prices should be monitored.

18.3 Monitoring prices at local markets

Especially in areas where official prices are not respected, data on prices should be gathered directly from the sellers. This can best be done by an enumerator who speaks the local language and lives in the area. While it is useful for the director to visit local markets and to see what is available, prices quoted to him or in his presence may be higher than the real prices.
In many places, women are involved in trade, usually for retail sales only but sometimes on a large scale. If the unit employs female enumerators, it may be desirable to have them gather data on prices for retail sales, with a male enumerator gathering data on bulk prices to traders.

18.3.1 Selection of markets

Since the data will be used mainly to evaluate the farmers' production, only those markets visited by the farmers need to be monitored. Those markets should first be identified by a survey on marketing habits. Frequently the farmers visit several markets around their village and sometimes go quite far away. A check should be made to see whether it is possible for the enumerator to reach the markets all year round. Distance is not necessarily a handicap, because enumerators and farmers are both likely to be willing to go out of their way to visit a busy market. The markets most popular with the farmers are then selected for survey. If there are too many, either those most important for the survey can be selected or simply a few at random.

While the main goal of a survey of marketing channels and prices is economic, it should be realized that markets in Africa are important social occasions and that farmers are likely to take the time to patronize several of them. This influences the time farmers consider available for farming, as well as the marketing channels available to them.

18.3.2 Introduction of the market survey

Before the start of the survey, the enumerator (preferably accompanied by the unit director or the supervisor) should pay a formal visit to the traditional chief and to the market chief of those markets selected for the survey to ask their permission for the survey. The reasons for the survey should be explained to them and it should be emphasized that the inquiries the enumerator will make have nothing to do with the tax office or the price control board. The market chief should then introduce the enumerator to the sellers, who will doubtless require a similar explanation.

The market chief can supply basic information about the market's periodicity and infrastructure, and may also be able to provide a list of the villages whose people visit the market.

18.3.3 Survey program for prices of produce

The price of agricultural produce, especially that of cereals, depends upon the quantities available for sale and the number of buyers, so prices tend to rise as the next agricultural season approaches. A monthly visit may not be sufficient to obtain a good estimate of seasonal fluctuations. Markets are held at regular intervals, sometimes once a week or, as in most of West Africa, every three days. Surveys conducted every twelve days (every fourth market) have been found to give satisfactory results. The frequency of visits should take distance and time into account, but in any case should be done regularly. The hour at which the survey is conducted is also important: it should take place at the busiest time of day when every type of produce sold is still available.

18.3.4 Survey form

To design the form for a market survey, the director needs a list of the various products that will have to be monitored. This can be found by examining what prices will be needed to evaluate farm results, and any specific requests from the project management. After visiting several markets and talking to farmers and sellers, the director may find that he can eliminate some products because they are not sold there or because gathering data on them is likely to cause difficulties.

Transactions at traditional markets are based on local weights and measures, usually identical for a few neigbouring markets. Some products, like peanuts and tomatoes, are sold at a fixed price; it is the quantity sold that fluctuates. The survey form should therefore include space for the price and the corresponding quantity so that the price per kilo can be calculated. The enumerator should be given some money to buy a few products each time so that he can weigh them later.
As prices are likely to vary between sellers and as the quality of some products will also vary, space for several price quotations for each product should be allowed on the form. Standards have to be set for the quality of the produce so that the prices can be aggregated correctly. For rice it is especially important to indicate the quality of the grain because this will influence the price. At the beginning of harvest, the enumerator should also note whether the price is for fresh produce or for last year's stock.
A form that covers several market days (e.g. for one month) makes it easier to check the data. As the monthly questionnaires are processed at the office, a table of average prices per kg per month for each product at each market can be kept up to date. If the prices at one market do

not follow the overall trend, this information can be checked. A graph of the time series for each product at each market makes it easy to notice any trend and takes only a few minutes to update each month.

An example of a survey form is presented in Table 18.1. It can be used for three subsequent surveys at the same market and has space for three price quotations for each product. It assumes that the units used for sale are fairly uniform (standard sack, tin measure). A similar survey form can be used to obtain prices for bulk sales to merchants.

18.3.5 Obtaining prices

Ideally, to avoid disturbing the sellers, the enumerator should walk around the market inquiring about prices without holding pen and paper, but should write down the information immediately afterwards at the edge of the market. He should inquire about the price of each product from several sellers, observing differences in quality. If the project is promoting an improved variety of a crop, he should note the prices of both the local and the improved variety. The enumerator should buy some items at each visit, as a good-will gesture towards the sellers and so that he can weigh the items at home afterwards. It is possible that the standard unit of measure is filled differently depending on the season. At the office, the average weight in kilograms of a local unit of measurement for each product can then be calculated and the seasonal price fluctuations evaluated. The information obtained by the enumerator should be written on the form exactly as he receives it: in local currency per local unit of measurement.
For goods sold in fixed quantities (e.g. firewood), the weight or volume should be measured in several sales from several merchants to find the average weight of the quantities in kilograms and the average price. The prices of some frequently used services (e.g. the price of grinding cereals at the mill) could be gathered in the same way.

To obtain the prices of small agricultural tools, the survey should be conducted when those tools are being purchased, and not after the end of the agricultural season.

Table 18.1. Example of a table for retail price survey of local markets (CFA per dish)

Name of market: .... Month: .... 19..

Produce	Date ...... Price (CFA)			Observations	Date ...... Price (CFA)			Observations	Date ...... Price (CFA)			Observations	Weight in kg per dish
	1	2	3		1	2	3		1	2	3		
White Sorghum													
Red Sorghum													
Millet													
Maize													
Sesame													
Soybean													
Rice (paddy) (first quality)													
(second quality)													
(third quality)													
Groundnuts													
Local Beans													
Cowpeas													
Remarks: (problems in survey, is market busy, and so on)													

The enumerator can check the accuracy of price quotations by listening when other transactions are taking place and by asking people he knows how much they paid for their purchases. As a rule, there will be variations in price between traders and at different times of the day. If an enumerator records the price of a product as remaining perfectly stable over a certain period, he may well be lying.
Data on prices for bulk sales to a merchant should also be obtained. One then has both retail and bulk prices available when calculating the value of the farmers' production.

18.4 Survey on prices of household goods

Prices of household goods (e.g. cooking pots and implements, lengths of cloth, batteries, tinned food) usually remain quite stable during the year. If price quotations on these goods are needed, a survey once or twice a year will suffice. These prices will be required in the following cases:

- If the monitoring unit needs to evaluate how much money an average household spends to cover its basic needs;
- If a survey of actual household expenditures is being conducted; knowing the usual prices of goods at the local markets helps in checking the accuracy of the data gathered from the farmers.

A survey on household goods need not be conducted early in the program. It will take time to know enough about the way of life in the households to determine their basic expenses.

19 DESCRIPTIVE DATA ON THE SAMPLES

Two categories of data will be needed for each sample monitored:

- The data asked for by the users when the program was being designed;
- Data describing the basic characteristics of the sample elements. This information provides a better understanding of the behaviour of the farmers and allows the sample to be grouped into homogeneous classes during analysis.

This chapter is concerned with descriptive data on two commonly used sample elements; the farm and the field.

Words like farm, field, family, and household can be understood to mean several different things. As used in this book, the words are defined as follows:

Household: Those people who live together in the same homestead under the authority of one person (head of the household). A household often includes several nuclear families.

Nuclear family: The smallest family unit in the household; it consists of a man with his wife (or wives) and their children.

Field: A piece of land controlled by one member of the household, on which crops are grown either in pure stand or mixed, or different crops are grown in pure stand side by side.

Farm: A farm can be defined in several ways, depending on the type of evaluation to be done.

- The aggregation of all fields which are 'owned' or leased by the people living in a household. This includes communal fields controlled by the head of the household and private fields controlled by individual members;

- The aggregation of all the communal fields directly under the control of the head of a household. Such a farm is regarded as one decision-making unit;
- The aggregation of all fields 'owned' or leased by the members of a nuclear family;
- The aggregation of all fields under the direct control of the head of a nuclear family;
- The aggregation of fields 'owned' or leased by one individual member of a household (private fields).

19.1 Descriptive data on the farm

Soon after the sample has been selected, the enumerator should visit each farmer chosen to obtain the descriptive data on the farm. Basically these only concern the size of the household, the distribution of its members over age classes, the number of workers, any additional characteristic used for stratification (to check whether the stratification is correct) and perhaps a list of the main crops grown. The goal of this survey is merely to ensure that the sample is acceptably representative.

Sometimes it happens that the information used to stratify the sample was incorrect and the sample is thus biased or has no stratification. It can still be changed at that time with no loss of data.
Once the sample has been checked, data are needed on the land available for cultivation and on the means available to the farmers to work it: labour, capital, equipment, and management capacity. These data are needed to place the performance of each farm in its proper context for analysis.
The survey of the farm's means of production will be conducted in one or more interviews during the early part of the first crop season. It should not be done too early, however, because much of the data desired is sensitive, so the enumerator should have had time to establish a good relationship with the farmers and to have become familiar with their situation. Also, the farmers are likely to be more cooperative when they have become familiar with the interviewing procedure and have got to know the enumerator personally.

Communal fields

19.1.1 Rights to land

In many areas of subsistence agriculture, land is not owned by individuals but by a group of families. All the descendants of those families who created a village have the right to cultivate some of its land. One descendant of the first family in the village (often called the 'chief of the earth') allocates fields to the farmers as needed. A farmer does not own the fields he cultivates; he merely has the right to use the land and to own the harvest.

Under such conditions, land cannot be sold, but the right to use a field can be inherited or 'loaned' to someone. The 'chief of the earth' can be a good source of general information about local traditions pertaining to land, the different kinds of land rights, and how land can be transferred from one household or individual to another. More specific information about who cultivates each field or part of a field and who is entitled to the harvest has to be obtained from the farmer himself. Part of the harvest probably has to be paid to the one who 'owns' the land, or some other form of 'gift' has to be made to him. This is a delicate matter because it reflects personal relations among the people.

Trying to differentiate between original right to use land and its temporary loan can be difficult. Collecting good data about the specific land tenure situation will not be easy.
The situation is made even more complex because two kinds of fields are usually cultivated by a household:

- Communal fields, which are cultivated jointly by all the workers of the household. The head of the household makes all management decisions and is entitled to the harvest;
- Individual fields, which are cultivated by those workers in the household who are entitled to their own piece of land. They work these fields in the time left over from their work in the communal fields. Each individual manages his own field and is entitled to its harvest.

Unless one wants to find out how income is distributed among individual members of the household, only the land transferred between households need be considered. Data on transfers of land within the family and the obligations they create are seldom required by the monitoring unit; such information would be difficult to obtain anyway.

19.1.2 Identification of fields

One of the first tasks of the enumerator will be to indentify the fields he is going to monitor. This can be quite tedious and complicated, as the enumerator and the farmer have to identify each field so that later they can be sure they are talking about the same one. They should go and look at the fields together and agree on the names by which the fields will be referred to.
A list should be made of all the fields of the household, with a description of the location of each field, what is being cultivated on it at that time, and any characteristic such as being located on low land. Even if income data are gathered for the household unit and not for each individual, the person who manages each field or part of the field should be identified. He or she will be the most knowledgeable source of information on that field and will be interviewed later for data on its inputs and outputs, its rotation, and any other details that might be required.

19.1.3 Size of the household

The size of each household and its composition can change from year to year and needs to be surveyed annually. There are various reasons for this survey:

- The number of people in a household is not a sufficient description of its labour force. A boy is not as strong as a man, and a woman will sometimes work fewer hours in the field because of her other activities;
- The household's potential labour force can be a major constraint on the amount of land it can cultivate and on the farming techniques it will use;
- When farm results are being analyzed, the number of people who have to be fed from the farm's production determines the level of production that can be considered sufficient;
- When the farm results of different-sized households are being analyzed, comparisons are only possible if the average production per worker and the average production per consumer are known.

Information on the households' size and composition should therefore include details on the sex and age of each of its members and whether he or she works on the farm or not.

Labour index

To allow the labour force of different-sized households to be compared, many projects use a labour index system, with the working capacity of an adult male as the standard unit and different indices for women, children, and old people. Working habits can vary greatly from one region to another and between ethnic and religious groups; this is especially true of the amount of work expected from women and children. So, if the monitoring unit decides to use a labour index, the choice of indices should be discussed with the project specialists, particularly with the social scientists.

Systems of labour indices suggested by FAO and in the *Memento de l'Agronome* are presented in Table 19.1.

Table 19.1. Labour indices by sex and age classes in West Africa

FAO			*Memento de l'Agronome*	
Sex	Age	Index	Category	Index
M*or F‡	11-15	0.25	Children	0.3-0.7
M or F	16-20	0.50		
M	21-60	1	M Adult	1
F	21-60	0.67	F Adult	0.5-1.0
M or F	over 60	0.50	Old people	0.3-0.7

* M = Male ‡ F = Female

Sources: FAO *Notions d'Economie Générale et d'Economie Rurale* 1973, 131, *Mémento de l'Agronome* 1980:1331

When all the individuals in a household have been indexed, these indices can be added up to obtain the household's labour index, which can be used in several ways:

- Households can be grouped into classes of labour index so that their practices and results can be compared;
- The amount of land cultivated by the household can be divided by its labour index to obtain the area cultivated per unit of labour. This can be a useful indicator for possible intensification of cultivation, while also allowing households to be grouped into classes of area per unit of labour for comparisons of their practices and results;
- The household's income can be divided by its labour index to calculate its income per unit of labour, thus making income data comparable between households.

The labour index is a most useful tool, but it has to be used with caution. Comparing results of projects that use different systems of labour index, for instance, will be meaningless and misleading. So too will be comparing results per individual worker with results per unit of labour. This will be illustrated by the following example.

Example of the use of labour indices

A household consists of four adults and five children, ranging in age from 2 to 62. In 1978, this household cultivated a total of 3.2 ha and had a net agricultural income of 86,000 CFA. Some ratios of this household, calculated with the systems of labour indices currently in use in two projects in the same West African country, are presented in Table 19.2.

Table 19.2. Labour indices in two projects

Members of household	Age	Labour index in Project A	Labour index in Project B
Man	62	0.50	1.00
Woman	58	0.25	0.50
Man	50	1.00	1.00
Woman	35	0.75	0.50
Girl	13	0.25	0.25
Girl	11	0	0.25
Boy	10	0	0.25
Boy	6	0	0
Girl	2	0	0
		2.75	3.75

If this household were in Project A, it would be given a labour index of 2.75, so its ratios (rounded off) would be:

Area per unit of labour: $\frac{3.2}{2.75} = 1.16$ ha

Income per unit of labour: $\frac{86,000}{2.75} = 31,300$ CFA

The same household in Project B would have a labour index of 3.75, so its results would be:

Area per unit of labour: $\frac{3.2}{3.75} = 0.85$ ha

Income per unit of labour: $\frac{86,000}{3.75} = 22,900$ CFA

The results obtained in both projects differ from the results that would be obtained from a simple division by the number of working individuals. Project A counts 5 workers while Project B counts 7, whereas in reality everyone except the two-year-old girl works on the farm.

19.1.4 Equipment

When monitoring the adoption of new farming techniques (e.g. animal traction), one has to find out what equipment is available to each household and what is its state of repair. The essential parts of the equipment should be mentioned: a plough is useless if the share is broken; farmers who have both a long and a short yoke are more likely to use their draught animals for ridging than those with only a short yoke. The condition of the animals is also important: two oxen do not necessarily make a pair if they are of very different sizes. Such obvious problems can be spotted by the enumerator, who should examine each piece of equipment himself. He should not attempt a detailed report of the working capacity of the animals unless he has received sufficient training.
The equipment inventory should be limited to what is really important for the unit to know. It is pointless to spend time ascertaining whether the family has eight or nine hoes; one can be sure that every worker able to weed will find a hoe available when he needs it.
An inventory of agricultural equipment is usually easy to make because most farmers realize that the equipment they own is an important factor in describing their farm. The aim of this inventory is only to make a rough estimate of the capital invested in the farm. Giving a residual monetary value to a used hoe or to 'farm buildings' such as a used chicken coop or granary does not make much sense. One only needs to know what is available and whether it fits the needs of the farmers.

19.2 Recapitulative form per farm

The descriptive data on farm units discussed in this chapter are needed for the classification of the farms and as a background for the analyses. As the data have to be updated each crop season, it is useful to group the main data for each farm on a recapitulative form as was explained in Chapter 8. As a reminder, the following suggestions are made:

- This form will group the basic information of one farm over several crop seasons. It will record the characteristics of the farm (farm size, household size, workers, inventories etc.), but can also be enlarged to include the crop results for each year. It should be on strong paper and should be filed so that it is easily accessible at all times;
- Prices used to value the crops may have to be changed over the years, so the form should provide space for the data in kind, as well as space for different price values.

19.3 Descriptive data on the field

In some surveys (e.g. those done to collect data on a certain crop), the sampling units are not farms but fields. The data required to describe each field are then obtained by interviewing the person who cultivates that field. The information should include:

- Location;
- Identification of person who controls the crop;
- Soil quality and topography, if relevant;
- An inventory of agricultural equipment available to the cultivator.

No data on the household are needed. Unlike data on farms, the data on fields should be kept on separate forms, one for each field, to make it easy to group the forms in different ways.

20 AGRICULTURAL PRACTICES

Data on agricultural practices have to be obtained in a number of separate surveys as will be explained in this chapter.

20.1 Previous agricultural experience

The attitude of the farmers towards proposed new techniques is greatly influenced by their previous experience with those techniques. Introducing animal traction to farmers who have already worked with animals is likely to have more rapid success than when introduced to farmers who have never worked with animals before.

To properly interpret the data obtained in farm monitoring therefore, one must be aware of the previous experience of the farmers. Details on this subject can be gathered in a single-visit survey. The timing of this survey is not important, although it should not be done right at the start of the monitoring; if done in the later stages of the first crop season, after the enumerator has gained the confidence of the farmers, it is likely to produce better results.

During this survey, the enumerator will gather the data by reading questions to the farmer. The questions need to be carefully worded to obtain correct answers. It should be perfectly clear, for instance, which questions pertain to the way things were done in the past and which pertain to the present.

As with any new survey, the survey on previous agricultural experience should be carefully introduced and explained to the farmers. They should be told that if the unit is to understand the data throughout the year, it must know how they farmed before and what changes the project has brought in their practices and results.

Two types of questions need to be asked:

- Questions to obtain a simple list of the key elements of the farmer's method of cultivation (crops grown, use of improved seeds, manure and fertilizer, animal traction, insecticides etc.);
- Questions asking for a comparison of the farmer's situation before and after the introduction of the project.

It can also be enlightening to ask the farmers what the project has brought into their lives which they would not otherwise have had. The items they mention and the order in which they mention them can show quite different priorities from those of the project management, and might help to clarify the attitude of the farmers towards the project.

To check the validity of this survey, one must know what the traditional farming system was. A talk with an extension agent who knew the area before the project started can be helpful here.

Traditional agriculture

If there are many open-ended questions in the survey, the results may have to be tabulated in two stages. First the data processors should go through the questionnaires, listing the answers found and identifying general categories. They can then do a full tabulation and count the frequencies of each answer. Grouping the farmers according to their previous experience with the techniques being promoted can show which practices are more readily adopted as the farmers were more familiar with them.

Sometimes a rough indication of the farmers' previous experience may already have been used to stratify the sample. If so, this survey can also be used to check whether the strata were correct. If found to be incorrect, a new sample may have to be selected - in addition to, or instead of, the first.

For example: A sample was selected with two strata: modern farmers (identified by the ownership of a plough) and traditional farmers (no plough). From the survey on previous agricultural experience, it has been found that many so-called modern farmers have never used their ploughs, whereas several farmers who were classified as traditional hired animal-drawn equipment to cultivate their land. The conclusion is that the stratification was incorrect, so another sample has to be drawn. This may be worthwhile only if monitoring is to be continued for several years. If no other sample can be selected, the analysis should take this stratification error into account.

20.2 The fields

The quality of the soil and the topography of a field influence that field's potential for crop production. If an analysis is to be made of the relationship between input and production for a sample of fields, a detailed description of each field is needed. The data required concern soil type, topography, crop rotation, and erosion protection. Even if the analysis is less demanding, some of these elements can help to explain the differences between the results obtained by different (groups of) farmers.

20.2.1 Soil type and topography

Information on soil can be difficult to obtain if a detailed pedological map is not available. If need be, the monitoring unit and the project's soil scientists can set up a simple classification system of soil type and topography to classify each field. This is worthwhile only if the soil scientists feel that there are significant differences between soil types within the project area and that those differences are likely to influence the performance of the farmer or his crop results.

Farmers usually have local names for the different soil types found in their region, and they know the type of soil of each of their fields. The classification system chosen by the soil scientists should be as close as possible to the local system because the local system is likely to be based on those factors that determine the choice of farming practices and the resulting crop yields in the area.
Only major elements in the topography need to be noted (e.g. whether the fields are on low lands, high lands, slopes etc.).

20.2.2 Erosion protection

Erosion can have different causes (mainly wind or rain) and therefore requires different protective measures. It is not the intention here to discuss erosion problems but only to underline the importance of including erosion-control aspects in a long-term survey. The negative effects of poor erosion control may only become visible after many years, but if a record has been kept of what has been done over the years, this can be useful.

20.2.3 Crop rotation

The crop rotation practised on a field (the sequence of crops grown on the field) can influence the potential of that field. In the long run, a good system of crop rotation can contribute much to maintaining soil fertility even if no fertilizer or manure is applied, whereas continuous cultivation of the same crop can exhaust the field and allow diseases and parasites to develop. Since past events are involved, the data on crop rotation can only be obtained by questioning the farmer. A major drawback in gathering data on the history of a field is that it is not always clear to the enumerator which field the farmer is talking about. A solution to this problem is for them to go to the field together and do the inquiry there. If the farmer has many fields and the crop rotation is desired for all of them, the survey can be time-consuming and confusing. For a farmer with many fields, it might be advisable to interview him several times, talking only about a few fields at a time.

The questionnaire for a crop rotation survey should be kept simple. A table with one column for each field is sufficient. Table 20.1 shows an example of a table that can be updated every year.

Table 20.1. Example of a crop rotation form

Name of field	....	....	....	Observations
Year	Crops cultivated	Crops cultivated	Crops cultivated	
1975				
1976				
1977				
1978				

If a crop rotation survey is to include such details as fertilizer use, ploughing in of crop residues, etc., extra columns can be added to the table. Usually, however, it will be difficult to obtain correct detailed information about past years, and it is better to start registering such details now and to limit data on past land use to whatever generalities one can obtain.

Checking the information obtained is almost impossible, but crops of the previous year can be checked by looking at the crop residue on the fields, while the rotation can be verified somewhat by comparing it with the cropping pattern actually found on the farm.

A survey to find out how different crops are rotated within a field is likely to be too complex to attempt because the boundaries within a field change over the years.

20.3 Agricultural inputs

Development projects usually introduce an intensified farming system with high input levels - a capital-intensive way of farming which means more risk for the farmer. If he uses no inputs other than seed from his own stock, the farmer stands to lose no more than the family labour input, which is a great waste but does not represent a money loss for him. If, however, he has bought an improved variety and fertilizer and the crop fails, he might end up with a negative cash balance in addition to the waste of family labour. It is therefore of the greatest importance that the results of the agricultural improvements being promoted be closely followed. A key element in this is the relation between input and production. The following sub-sections will describe how to gather input data through questionnaires and how these data can be checked through observations in the fields.

20.3.1 Units of measurement

The units of measurement used by the farmers (e.g. the amount of grains contained in a basket or a dish) can vary from one region to another. The basic data recorded by the enumerators should be in those local units of measurement, however, because in this way the validity of the data will be greater than if the enumerator first has to convert the local measurements into internationally used units. A separate inquiry can evaluate the local units of measurement in terms of international systems so that data from different regions can be compared.

20.3.2 Seed

The quantity of seed used by the farmers to sow their fields is difficult to monitor, because farmers sometimes take the seed out of their granary, take more seed for a second sowing, if necessary, and use whatever seed is left in the granary for food. The enumerator will rarely be able to weigh the amount of seed used, but can estimate it by taking average quantities of seed needed to sow one hectare as shown in Table 20.2. He should endeavour to measure the seed used by a few farmers to check whether these estimates are close to reality.

Table 20.2. Quantities of seed (in kg) required to sow one hectare

	Maize	Sorghum	Millet	Rice	Cotton
Seed quantity	15-25	4-8	5-12	30-100	15-40

If credit is given to buy improved seed, it is easy to find out how much each farmer in the sample has bought. This will usually be close enough to the real input. If the project is introducing improved varieties, information is also needed about the variety used. It must also be known whether the seed is sown in rows or broadcast, and whether it is protected with insecticides. These factors have to be known when one is determining the yields obtained, and will be used to classify the yield data for the analysis of the relation between input and agricultural production.
The period in which the sowing (and any resowing) takes place should be noted as it will be needed when the results achieved are being evaluated.

20.3.3 Fertilizer

High-yielding varieties only reach their full potential if they are properly fertilized. Fertilizers can be either organic (manure) or chemical. If the animals are grazing freely, only the house compound will benefit from their manure because the animals are corralled there at night. Even if a farmer uses manure on a large scale, it will still be difficult to evaluate the quantity he uses and its nutritive value; one has to be satisfied with only a rough indication.

Chemical fertilizer is easier to monitor since it has to be bought, either at the extension warehouse, sometimes on credit, or with cash at the local market. Often, however, for a variety of reasons, the quantity of fertilizer really used is less than the quantity bought. Correct data can only be obtained direct from the farmer - and only then if a good relationship exists between the farmer and the enumerator. During the period of fertilizer application, the enumerator should regularly ask the farmer whether he is using fertilizer and how much of which kind he is putting on each crop (in the farmer's own unit of measurement, which will probably be in terms of the sacks in which the fertilizer is sold at the extension warehouse: 1/2 sack, 1/4 sack etc.).
If the farmer has bought his fertilizer at the local market, he will express the amount he uses in terms of the market unit of measurement. This, together with figures on the price he paid, can give the enumerator enough information to evaluate the quantity and cost of the fertilizer used by the farmer.

The quantity and kind of fertilizer used are important aspects of the fertilizer input but are not the only ones: two others are the methods of application and its timing. Fertilizer application at the foot of the plant, between the rows, or broadcast before or after sowing all influence the plant growth differently. The economic benefit of fertilizer decreases if it is applied too late, whereas an early application implies a risk of fertilizing the weeds only.

Besides the factors which are under the control of the farmer, many other factors influence the effect of fertilizer. One of these is the soil moisture. Fertilizer applied when the soil is dry has hardly any effect. Soil moisture can be measured directly, but usually rain and evaporation figures are sufficient.

20.3.4 Insecticides

Some crops (e.g. cotton, cowpeas) can suffer severely from insect infestation and have to be sprayed regularly. The quantity needed to spray a hectare effectively is known, depending, of course, on the type of insecticide used. The number of treatments required depends on the insect population. For an evaluation of crop results, one must know whether the right dose of insecticide is applied each time. Slight overdoses of insecticide do not matter but a low dose does not kill enough insects. The farmer can be asked how many cans of insecticide he used. To evaluate whether the treatments were frequent and timely enough, the project's agronomist can be asked to make regular checks on the insect population. For crops like cowpeas, the farmers may not want to use chemical insecticides because traditionally the leaves are cooked and eaten.

20.3.5 Checking the data

Input data are not gathered at regular intervals throughout the season since activities like sowing and fertilization are time-specific. At the appropriate time therefore, the enumerator should ask only those questions about inputs which are relevant at that stage of the agricultural season. As far as possible he should check the data by observations in the field. Frequent visits to the fields to see what activities are taking place and what inputs the farmer is using will greatly help the enumerator in his interviews. He will know which fields are likely to be mentioned and can prompt the farmer if something is forgotten.

Another effective way of checking input data is through the agricultural credit unit. Modern inputs are often marketed via, or in cooperation with, a credit unit, so it can supply information on the quantity of

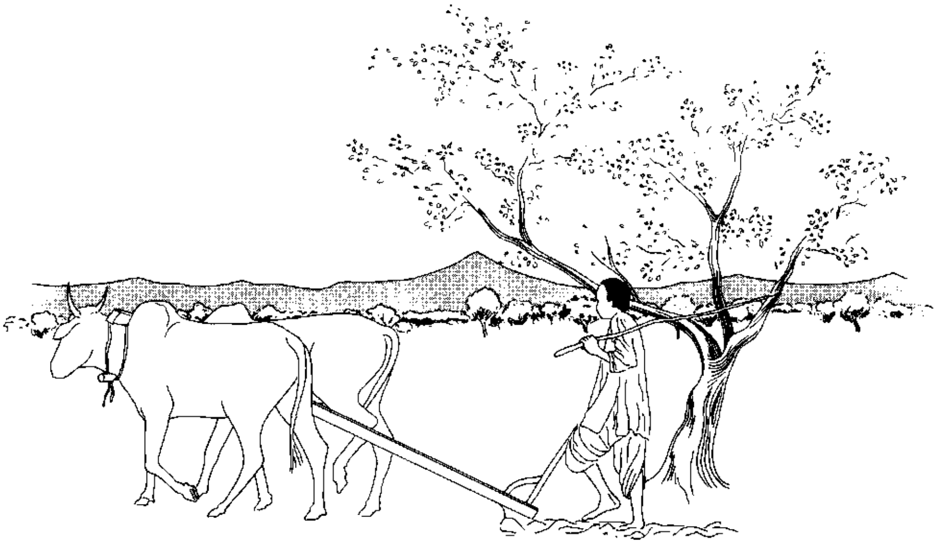
inputs actually received by the farmer. This information should be used with care, however, because farmers sometimes sell some of their inputs at the local market if they are short of cash, or may store some for next year if supplies have proved to be uncertain, or if the credit unit obliges them to buy a certain minimum quantity. Conversely, they may buy additional inputs at the local market.
Sometimes, a farmer does not remember the quantity he has used, but might be willing to show what is left over from the quantity he has bought through the credit unit.

20.4 Cultivation practices

Inputs alone are not sufficient to explain the results achieved. These depend also on the cultivation practices and their timing. Although it is difficult to obtain quantitative information about cultivation practices, it is fairly simple to register some key factors: type of land preparation, timing of sowing, and timing and thoroughness of weeding.

20.4.1 Land preparation

The way in which the land is prepared before sowing determines the quality of the seedbed and the soil's moisture storage capacity. In traditional farming, land preparation is limited to removing the weeds. A more thorough land preparation involves cutting and turning the top soil by ploughing, and breaking up clumps by harrowing. Both these processes require animal traction.
The quality of land preparation depends upon its timing and depth. Preparing the land too early in the season, when the soil is still very dry, can lead to wind erosion. The depth of land preparation should be adapted to local soil conditions. It can only be done well if the equipment used is in good condition and is properly adjusted. A well-adjusted plough is easier to work with and the depth of ploughing is uniform. When new equipment is being introduced, one of the main tasks of the extension service is to teach the farmers how to adjust and maintain it.



Farmer using animal traction

20.4.2 Sowing

The time of sowing can have a great bearing on the yield, as can the method of sowing - broadcast or in rows - because of the resulting plant density and ease of weeding. The tools used for sowing do not affect the yield; a seeder does not necessarily give better results than sowing by hand; it only makes it possible to sow faster.
Deciding when to start sowing is one of the major decisions a farmer has to make - and a difficult one - because he can never know in advance when the rains will be frequent enough to ensure good plant growth. Once the rains are well established, all fields must be sown as quickly as possible.

20.4.3 Crop maintenance

Three aspects of crop maintenance need to be monitored:
- Its timing in relation to sowing;

- The techniques used;
- The equipment used.

The timing of weeding can affect yields because the longer the farmer waits after sowing, the longer the weeds grow in competition with the crop and the lower the yield will be. Nor should thinning be done too late; if too many plants grow in the same place for too long, the plants that remain after thinning will remain smaller than they would have if thinning had been done on time. On the other hand, thinning too early is risky because the young plants are still very sensitive. But the timing of these activities can only be interpreted in relation to field conditions: later weeding does not matter if there are few weeds, and weeding on time simply means keeping the field clean. It can be useful to give the enumerator example pictures showing how to classify different stages of weed invasion.

The technique of crop maintenance can consist of a simple weeding by hand or with animal traction, or it can include ridging (which is usually done with animal traction but can be done by hand). Ridging helps retain soil moisture and lowers the risk of erosion.

The type of equipment used for crop maintenance will affect the quality of the cultivation. The right tools have to be used to achieve good results and they have to be well maintained and properly adjusted.

20.5 Labour inputs

New farming techniques require the farmer not only to invest money in his crops, but also to follow a fairly precise agricultural calendar. A higher yielding variety is often more sensitive to late sowing or to poor weeding. (The local variety may yield better under poor weeding.) Fertilizer, which accelerates the growth of weeds as well as that of the crop, makes early weeding more important than ever. For these reasons, labour availability influences the farming practices that can be adopted. Most monitoring units will therefore be asked to do some kind of labour survey.

A labour survey is difficult to do. Each farm has many fields and the members of the household work together on some fields and alone on others. Few farmers have a watch so estimating the labour input in hours - a unit unknown to many of them - will be almost impossible. But they do use the position of the sun to mark the time of day, and Moslem prayers also provide reference points. So, depending on the objective of the labour survey and the level of precision required of the data, it may be possible to select a methodology that provides enough information to meet the needs. A few possible methodologies are discussed in the following sub-sections.

20.5.1 Identification of labour bottlenecks

Doing a labour survey is only worthwhile if the availability of labour is a constraint. A reconnaissance labour survey will identify any labour bottlenecks that might exist at certain times and might hamper the adoption of more labour-intensive farming techniques. Such a survey can be done in several ways.
A rough approach is to conduct a single-visit survey to gather data about farm size and the number of workers on each farm. These data can be used to calculate the man/land ratio, which can then be compared with the theoretical labour requirements of the crops grown in the area. This will identify the periods when labour bottlenecks are likely to occur. Precise data on farm size, however, are not always easy to obtain.

Another way is to monitor the labour input on all the fields of a sample of farms in more detail, but without trying to quantify the labour input. The enumerator observes the fields and the performance of the crops on those fields. He talks regularly to the farmers, inquiring whether farming activities are taking place as planned. If a farmer is late in relation to his neighbours, or if he says that he is behind schedule, the enumerator should try to find out why. This can be a very sensitive subject, so the enumerator should wait until the farmer himself mentions the matter. The enumerator should never be first to mention it because that might sound as if he were passing judgement on the farmer, who is then likely to say that circumstances beyond his control prevented him from working more. Instead, it is wiser for the enumerator to ask the farmer what he plans to do in the period between now and the enumerator's next visit, which should be not more than a week later.

The data from field observations and interviews should be recorded on the same questionnaire. A table is probably the most practical format. Ample space should be left to note the reasons given by the farmers for being behind schedule, which the enumerator should note in as much detail as possible.
A labour survey like this one provides a qualitative understanding of the labour situation and of the factors restricting the labour supply of the households. This information can be useful in planning the introduction of new farm techniques, but is inadequate for planning the cropping patterns of the farms.

20.5.2 Detailed labour surveys

A more detailed labour survey is needed if labour inputs have to be quantified for different crops. A labour film for each crop (i.e. a summary of the labour inputs on a crop over time) is needed when planning optimum cropping patterns and optimum areas to be put under cultivation by households of different sizes.
Before starting a detailed labour survey, one should understand that the labour input of a household varies each year - with the rainfall pattern, the performance of the crop, and the timing of operations. The survey should therefore continue over several crop seasons. The 'average' labour film obtained in this way is a good indicator of the amount of labour the farmers put into each crop at different periods of the season. Taking existing constraints into account as it does, it may not correspond to the labour requirements for crops given in agronomy books!

A detailed labour survey can be done in two ways:
- Using a day as the unit of measurement;
- Using an hour as the unit of measurement.

Labour survey in days

A labour survey using days as the unit of measurement will be found sufficient for most purposes. If need be, the day can be split up into four quarters: two in the morning and two in the afternoon. The survey is done by a combination of interviews and observations on a sample which must be kept small. As the farmer does not keep records and has to rely on his memory to provide information for all the workers in his household, he has to be interviewed often, say twice a week. The enumerator could divide the sample into two groups, interviewing one group, say, on Mondays and Thursdays, and the other group on Tuesdays and Fridays. On those days, he inquires about the work done by the farmer and the members of his household during the three or four days since the last inquiry, starting with the day before the interview, then two days before, and so on, reviewing each field in turn. Some farmers may find it easier to review the work of each individual in turn.

When not engaged in interviews, the enumerator will be making observations in the field, noting who was seen working, what kind of work they were doing, and the problems they encountered during the work. Knowing these facts when he interviews the farmers, he can check that what he has observed during the last three or four days coincides with the answers they give; he will also be able to prompt the farmers if they are unable to reconstruct the situation in their minds.
If the enumerator records his observations on a preprinted table or form, there will be less chance of his forgetting anything. An example is given in Table 20.3.

Table 20.3. Example of a table to record observations during field visits

Date and time of observation	Name of field	Type of work	Number of workers				Problems encountered
			Age class				
			0-<10	10-<15	15-<55	55+	
			M F	M F	M F	M F	

The date and time of observation (early morning, mid-morning, etc.) and the name of the field are also needed for administrative purposes, to check whether the enumerator is meeting the entire sample regularly and whether all the fields are being observed.
The information obtained about the type of work being done, together with an indication of the approximate age and sex of the person doing it will not only be helpful to the enumerator in checking the data obtained during interviews, it will also allow labour indices to be used later. These will make it possible to compare the labour input of different households.
Noting any problems encountered in the work being done will help explain any anomalies in the data. An inexperienced farmer working with animal traction, for instance, may be taking an overly long time to plough his field.

The survey forms or tables that the enumerator fills in during interviews can also be preprinted. An example is given in Table 20.4. In this table, the day is divided into four quarters. If more than one operation is done in the same quarter day, only the one that took the most time need be noted. If a member of the household worked on more than one field during the day, a separate line should be used for each field. The column for observations can be used for any additional information that might help to clarify the data.
The enumerator should work to a strict schedule because labour surveys cannot be analyzed if they are incomplete. If delays have occurred, the enumerator should be given some leeway to enable him to catch up. For this, he may need a special form to gather the missing data.

After the survey forms have been brought back to the office, processing should be done with great care. Aggregating the data into a labour film per crop is complex because the labour days of men and women and of people of different ages are not equal. For this reason, the labour input of each has to be weighed carefully against the labour indices attributed to them (Chapter 19).

Table 20.4. Example of a table for one day of a labour survey

Name of the farmer: ....

Village: .... Date: ....

Name of worker	Sex	Age class	Field	Crop	Type of work	Part of the day*				Observations
						Morning		Afternoon		
						I	II	I	II	
From household:										
Exchange of labour:										
Paid labour:										

* Fill out M if work is done by hand during that part of the day

D if work is done with donkeys during that part of the day

O if work is done with oxen during that part of the day

\- no work is done during that part of the day

Labour survey in hours

If a detailed survey of labour is needed, using hours as the units of measurement, it can only be accurate if the enumerator surveys just a very few farmers or a group of fields located alongside one another. Only in this way will he be able to observe the time of arrival and departure of all workers, and their periods of rest. Merely asking the farmer about hours will result in a large margin of error.

The work of the enumerator is simple: he spends the entire day in the fields, watching the farmers and noting any additional information that might provide a better understanding of the survey results.

When aggregating and interpreting the results of the survey, the different agricultural activities have to be kept separate. A farmer's work capacity in hours per day depends on the operation undertaken: an hour spent on weeding takes more energy than an hour spent on sowing. This has to be taken into account when the labour film is being put together.

Detailed labour data are difficult to obtain and relatively expensive to gather because the survey immobilizes the enumerator throughout its duration. Considering the time and money required for such a precise survey - especially if a representative group of farmers has to be surveyed - it is worth considering whether such precision is really required before starting the survey. The results of such precise labour surveys are rarely used in their entirety, which forces the conclusion that a less detailed labour survey will fulfil the needs of the users at a much lower cost.

21 ESTIMATING AGRICULTURAL PRODUCTION

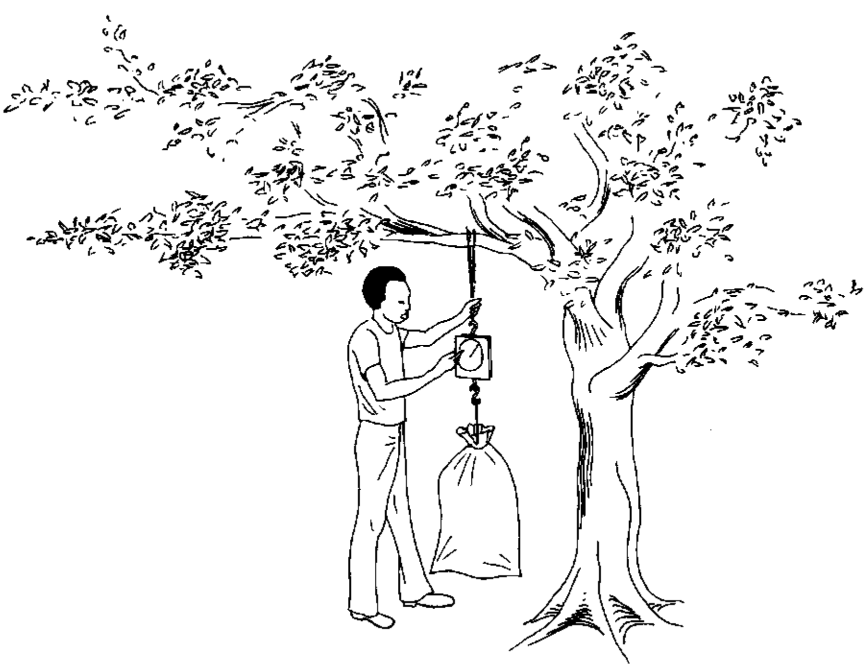
Although the income derived from the cultivation of crops does not represent the entire income of subsistence households, technological improvements are most often aimed at raising crop production. When a monitoring unit needs to measure the results of technological change at the farm level, it therefore needs estimates of crop production in kilograms or of yields in kilograms per hectare. There are many methods that can be used to make such estimates. The decision on which method to use depends on three things:

- What is the user going to do with the information? If he intends to make an economic evaluation of farm results, he will need total production figures only. If he wants to evaluate the performance of new seed varieties, he will require yield measurements. If both analyses are to be done on one sample, two of the three aspects of production (total production of the field, yield, area cultivated) can be measured, in whichever combination happens to be the most practical;
- What type of farming is practised by the farmers? A field cultivated traditionally requires a different method from a field cultivated in rows;
- What crop is grown? A good method for cotton is not necessarily a good method for cereals.

In this chapter some simple methods of estimating the three components of agricultural production will be presented. The problems likely to occur during data gathering will also be discussed.

21.1 Yield estimate by sample plot

A yield estimate by sample plot is done by picketing off a small part of the field, harvesting it separately from the rest, and calculating how much would have been harvested on one hectare. If the area of the field is measured as well, the total production of the field can be calculated. Alternatively, if the total production of the field is known, it can be used, with the yield of the sample plot, to estimate the area of the field. This system can be applied for every crop, whether grown under traditional or modern agriculture and in single or mixed cropping systems. It has the advantage that yield estimates can be done directly by the enumerator, with only minimum participation by the farmer, making it more likely that valid data will be obtained.



Enumerator weighing the harvest of the sample plot

The enumerator can picket off the sample plots at any time during the crop season, as long as he has finished doing so before the harvest begins. When the crop is still low, however, it is easier to place the sample plots, and the enumerator will be less influenced by the performance of the crop in choosing the location of the plots. The sample plots should be identified systematically and their boundaries clearly marked. The enumerator will need to check regularly whether the pickets are still in place.

The number and size of the sample plots in a field will determine the reliability of the yield estimate. The larger the plots and the more there are, the more reliable the estimate will be. In deciding on the number of sample plots, one must take into account the working capacity of the enumerator and the heterogeneity of the crop. The harvesting and threshing methods used on the sample plots should be the same as those used by the farmers.

21.1.1 Sample plots in traditional fields

In traditional fields, the farmers grow one or more crops irregularly spaced on a field of irregular shape. Sample plots on such fields have to be placed so that they represent the field correctly. If several sample plots are to be placed, their location can be chosen in two ways, either at random (which is the easiest way) or systematically so that all the plots are spread regularly over the field. In practice, the enumerator can rarely place more than one or two sample plots in a field, so they might just as well be placed at random. Whatever is done, it should never be left to the enumerator to select 'representative' parts of a field, because he may be biased in his choice.

Of the many methods of selecting a sample plot at random, one will be discussed in detail. The procedure is as follows:

- Accompanied by the farmer, go to the field where the sample plot is to be placed, taking with you a table of random numbers, a measuring tape, and some pickets. (The farmer can be asked to cut the pickets beforehand for an agreed fee.);
- Determine the longest diagonal of the field. This can be done visually, as there are no serious consequences if the wrong diagonal is chosen. Select any point in the random table as a starting point and read the numbers on the same line, in the same column, or diagonally in sequence. (See Chapter 17 for instructions on how to use a table

of random numbers.);

- Starting at one end of the longest diagonal of the field, walk along it as many steps as are read in the random table, using three digits if the diagonal is estimated to be longer than 100 steps. Otherwise use two digits. If the number read in the table is larger than the diagonal is long, take the next two digits (or three, depending on the length of the diagonal);
- From that point on the diagonal, walk as many steps perpendicular to the diagonal as the random table shows (next two digits). (When a series of sample plots are being placed in a series of fields, walk to the right of the diagonal in the first field, to the left in the second field, and so on alternately.);
- Start measuring out the sample plot from that point on, marking the corners with pickets. (Care should be taken to place the pickets deep enough in the ground that they cannot be easily dislodged.) It does not matter how the sample plot is placed from that point as long as it is done in the same way for all the fields. For very small sample plots, a wooden or metal frame can facilitate measurements, although the sample plot usually applied is a square of 10 × 10 m - too large for such a frame. Check whether the square is really square by measuring its diagonal, which should be about 14 m.

The square sample plot of 10 m × 10 m will be sufficient in most cases, but can be replaced by several smaller plots if the field conditions or the crop stand are irregular. A sample plot that includes the edge of the field should be avoided because the edges of a field are seldom representative of the field.

Example of a sample plot in a traditional field

An enumerator is to place a sample plot in the irregularly shaped field shown in Figure 21.1. The longest diagonal is AB. The enumerator is at A and will start pacing from there. This is not the first field he does; he has been reading from a horizontal line in the random table (Table 21.1) and is now at Line 14 Column 4, first digit. Pacing 414 steps is too long for the field so the next number (078) is used and the enumerator arrives as X_1. The next two-digit

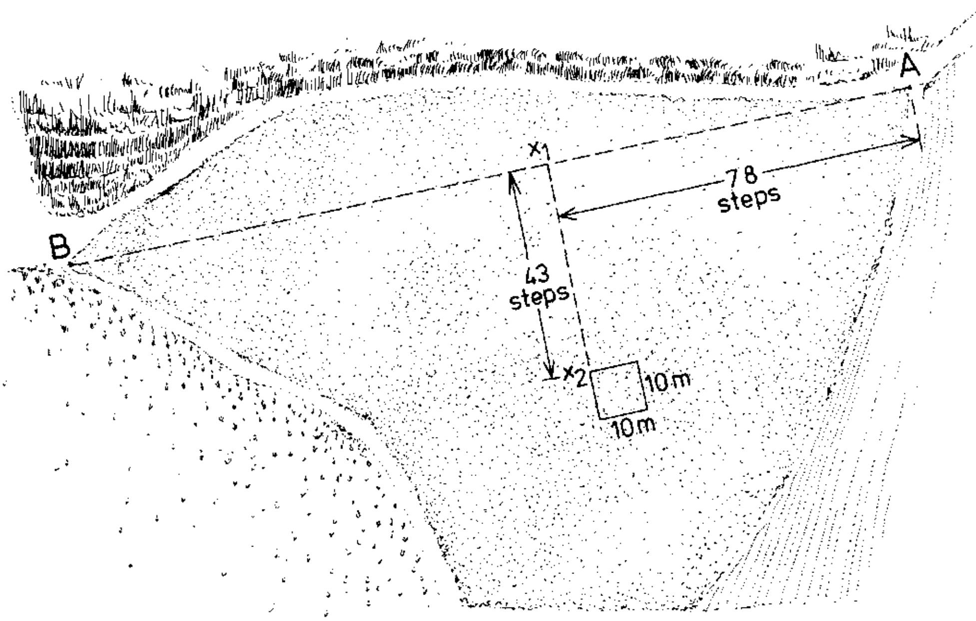


Figure 21.1. A sample plot in a traditional field

Table 21.1. Random numbers

	1	2	3	4	5	6	7	8	9	10
1	35994	98850	60645	47230	11263	99741	03582	30302	19643	77763
2	35344	66337	23823	72763	22369	99807	70900	55238	83395	15261
3	13823	62839	94098	55189	95052	37627	18776	00487	56759	50748
4	98609	33844	18419	46382	21125	36269	44139	82930	88609	68568
5	85899	92331	16951	47097	50041	80599	65130	99875	26349	81152
6	29189	54644	98913	51947	75933	15507	35643	89179	42506	70953
7	05702	91196	67141	07335	94480	21070	74326	54236	75337	25949
8	91229	76151	43632	08102	13994	15529	56338	61738	64360	79962
9	74455	84531	20710	21288	17400	26916	23756	14270	25379	13970
10	88082	84419	02310	31569	95456	08487	61218	19243	14704	83644
11	37650	84533	94066	82425	68634	87165	91454	67924	89277	51246
12	70941	32076	11270	30482	50987	12250	16421	26277	49515	83739
13	37089	92630	79234	94344	58856	19405	17165	90585	58610	62884
14	55793	45521	79262	41407	89092	43135	94170	97859	11158	70937
15	12016	83870	37223	16277	58520	63383	59563	03050	31501	02558
16	74210	26306	76189	94150	78056	45999	10514	53639	10769	14270
17	82612	66581	88352	94158	81060	71195	20366	83292	95216	49271
18	24686	48363	80991	88500	97464	04812	96211	99007	45035	32276
19	55425	41820	76775	47193	81795	05076	46937	60088	87967	53353
20	91748	41776	06702	59071	40726	83102	56119	63047	33845	29542

From: *Inleiding tot de Statistiek*. H. Rijken van Olst. Van Gorcum & Comp. N.V. Assen, 1966.

number (90) is also too large, as is 92, but 43 (third two-digit number) can be used. The sample is taken to the right of the diagonal and the corner of the sample plot is found at X_2. From X_2, the sample plot of 10 m × 10 m is measured out and the corners are marked by pickets. For the next sample, he continues on Line 14 Column 6 third digit of the random table.

21.1.2 Sample rows or plots in non-traditional fields

Most agricultural projects introduce some kind of mechanization at farm level, which demands that the farmers grow their crops in rows. This makes yield estimates by sample plots easier to do. Instead of placing a square plot in the field, a few rows (or parts of rows) can be chosen to represent the field. These rows can be distributed over the field to represent any irregularities in the field.
If entire rows are taken as samples, they can be marked by placing pickets at the beginning of the row. If only parts of rows are taken, a picket should be placed outside the rows to indicate those with sample sections, while in the rows themselves, the beginning and the end of the samples are marked by pickets. Measuring out sections in rows can be fairly troublesome and in general it is simpler to take complete rows.

As the fields will seldom be rectangular, the sample will vary from one field to the next. The area being sampled can be calculated by measuring the length of the row and multiplying it by the spacing between rows. This can be found by making several measurements of the distance between the sample row and the next row (on either side) and taking the average.

It is also possible to apply the same system of sample plot selection as was described for traditional fields. In choosing the location of the sample plot, however, one must take care that the rows of the crop are parallel to the diagonal of the sample square. Placing the sample plot in this way means that the number of plants in the sample will be about the same no matter where the plot is located in the field. Placing the plot with the sides of the square parallel to the rows of the crop could mean that shifting the plot by only a few centimetres will lead to the inclusion (or exclusion) of an entire row of plants. Yield estimates obtained from such a plot would not be accurate.

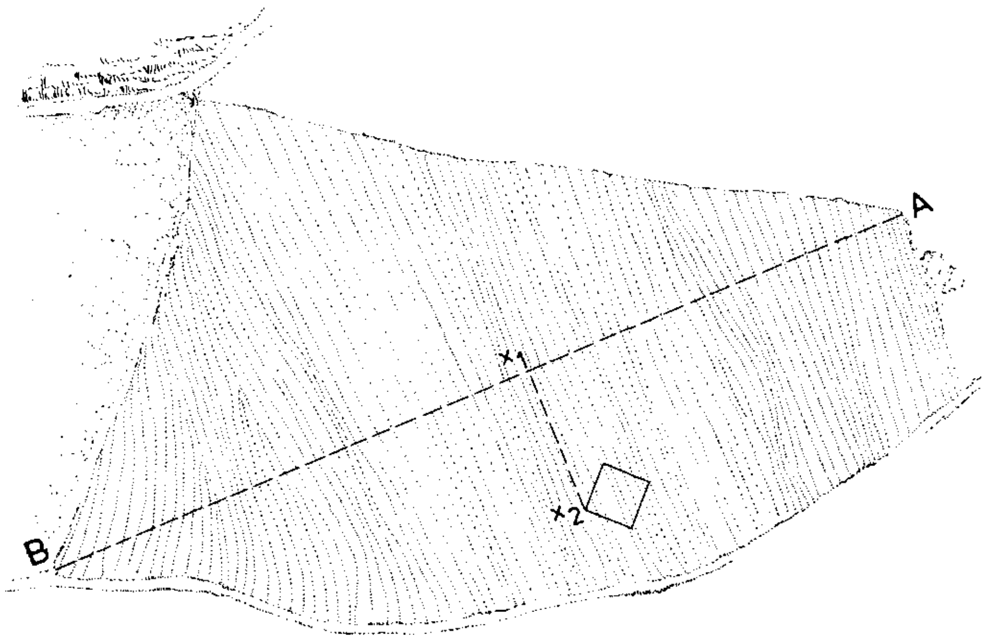


Figure 21.2. A sample plot in a field sown in rows

21.1.3 Harvesting the sample plots

The sample area should be harvested, preferably by the enumerator or otherwise by the farmer with the enumerator present, at about the same time as the rest of the field is being harvested. Early cutting is possible but correction factors then have to be introduced, so this should be avoided if possible.

If the farmer starts harvesting the rest of the field before the sample is harvested, he should be asked to leave an area untouched around the sample to be sure that no part of the sample is harvested by accident. This is especially important when harvesting is being done by children, who can easily spoil the samples. If the farmer harvests the sample, he should keep it separate from the rest of the harvest in sacks he has received from the enumerator for that purpose. The enumerator should

check that this is done correctly.
In the case of mixed cropping both crops have to be evaluated separately and if possible the harvested area of both crops determined. This will be easy in most non-traditional fields, where mixed cropping is usually done systematically (e.g. in alternate rows). In traditional mixed cropping, however, it will be difficult, if not impossible, and the crop combination will have to be regarded as one unit.

To evaluate the weight of the sample harvest, the easiest way would be for the enumerator to take it to his house and have it dried and weighed there. However, unless funds are available to buy the sample harvest from the farmer, this is not advisable because it will always lead to contention; the farmer will claim that he did not get his own crop back, or not all of it, etc. This would greatly damage the relationship between the enumerator and the farmer, so the sample harvest should stay with the farmer to be dried and weighed. If the farmer is assured that he can keep the sample harvest, he will be less likely to deceive the enumerator at some point during the process.

The weight of the sample harvest should be recorded on a form. If the yield is calculated in kilograms per hectare, the results can be compared with those obtained by other farmers. If the area of the field is known, the enumerator can calculate the total production of the field. As the farmer is very interested in knowing his total production, his relationship with the enumerator will be strengthened if he is given this information.

21.1.4 Sources of error in yield estimates

Yield estimates are subject to many errors which cannot all be avoided. If one is aware of the different sources of error, however, one can organize the data gathering in such a way that errors are kept to a minimum. The most serious sources of error are the following:
- Missing data because pickets marking the sample plots have disappeared. A check just before harvest can avoid this problem because the boundaries of sample plots can be identified again if need be;

- Errors in measuring the sample plot. These can largely be avoided by giving the enumerators a thorough training in measurement techniques. The supervisor should check some of the sample plots at random. Those that do not meet an acceptable standard have to be done again;
- Errors due to the location of the sample plot. In a heterogeneous field, one sample plot will not represent the field correctly. More sample plots can correct this situation;
- Errors in harvesting the sample. Sometimes the sample plot has already been partly harvested by the farmer when the enumerator arrives. The harvested production then represents an area smaller than the sample plot. Such errors can only be avoided if the enumerator has a good relationship with the farmer, maintaining regular contact with him and being on the spot to help him when necessary;
- Errors in weighing the harvest. This can be avoided by training the enumerator to use scales correctly and by choosing a very simple scale with widely-spaced numbers that are easy to read.

All events or conditions that can help to explain the yield and correct any errors (rocky field, flooded area, damage by animals etc.) should be observed and recorded by the enumerator.

21.2 Estimates of total production

It is not always possible to estimate yields from sample plots. Sometimes, in very tall crops such as cereals, the pickets cannot be found. Other times the area cultivated is not known so that the total production - an important variable in most farm monitoring programs - cannot be estimated by a crop sample. In such cases, the total production of the field or its area has to be measured. A simple procedure for direct estimates of total production will be discussed in this section. This procedure depends greatly on the cooperation of the farmer, making a good relationship with him more important than ever.

After the crop is cut on the field, the farmer transports it to his homestead and, using baskets, transfers it to his granary. There are then several ways of obtaining information about the quantity he has harvested, depending partly on the way the farmer is used to handling his crop.

21.2.1 Basket method

The basic idea of the basket method is to count the number of basket fillings harvested from a field. The farmer has to cooperate by using only one type of basket to fill up his granary. Ideally, the farmer fills up the granary in the presence of the enumerator and both count the number of times the baskets are filled. Some Moslem farmers are used to counting their harvest in this way to calculate the alms they owe. To find the average weight of the crop contained in the baskets, the enumerator takes a full basket once in a while and weighs its contents. If the harvest is stored unthreshed, he should thresh the contents of a few baskets and find their average weight in grains.

This method can work well if the enumerator has regular contacts with the farmer and arranges with him to put the entire harvest into the granary in one operation. To avoid inaccuracy, it should be done as early as possible before the farmer has sold or consumed part of his harvest. Early consumption of the crop can bias this kind of yield estimate. If the last year's harvest was not sufficient or if the crop is preferably eaten fresh (e.g. maize), early consumption can make this method unsuitable. The production of crops grown close to the homestead is especially difficult to estimate because the members of the household often consume part of the harvest a bit at a time before they start harvesting systematically.

The enumerator can evaluate the part already consumed by asking the farmer to monitor the quantity consumed, although for crops that are heavily consumed before harvest, it might be better not to use this method. It is important to watch carefully that the farmer does not mix the production of different fields.

21.2.2 Evaluation of the granary

A more difficult method, but the only one that can be used if the enumerator arrives after the harvest has been put in the granary, is to evaluate the contents of the granary. This can be done by measuring the

dimensions of that part of the granary which is filled up. In this way a rough estimate of the total harvest (often from different fields) can be made. The part of production that has been consumed will be difficult to monitor and the same applies to the part sold, which can be quite large, since most sales occur just after the harvest.
This method can also be used to check production data that have been gathered but are considered doubtful.

21.2.3 Errors in total production estimates

Data on total production are subject to error on the part of the farmer, who might not tell the truth about his production. It can be a very sensitive subject. Fear of taxes or forced marketing might cause him to lower the production data, and so too might the existence of a food aid program for which he wants to be eligible. On the other hand, he might raise his production data if a credit program is going to be started and he wants to be eligible for that. Total production data therefore have to be used with caution.

Apart from incorrect information from the farmer, other possible sources of error are:

- Weighing errors, as mentioned in 21.1.4;
- The farmer might mix up the harvests of different fields. This can be avoided if the system of estimating the production is carefully explained to him;
- Errors in counting the baskets. Especially if the farmer puts his production in the granary a bit at a time, it is easy to lose track of the number of baskets harvested;
- Sometimes the farmer stores his crop in different ways, depending on the quality of the produce, separating the best part from the rest. The weight of an average basket should therefore be found before that separation takes place; otherwise each part has to be treated separately;
- As mentioned before, that part of the crop already consumed is difficult to evaluate, and this source of error remains difficult to eliminate.

21.3 Measurement of a field

The fields cultivated by the farmers in semi-arid West Africa have irregular shapes and their area is difficult to measure. Measuring such fields is time-consuming and is subject to many errors, especially if the enumerator is not well trained for the job. It should therefore first be decided whether it is really necessary to know the area of the fields. If so, it must then be decided how precise the measurements have to be so that the most efficient method of measuring can be chosen. A few such methods will be discussed here, chosen for their simplicity and because they do not require complicated equipment. The shifting cultivation practices of the farmers will mean that new measurements have to be made every year.

21.3.1 Measuring on documents

Maps showing the boundaries of fields can be made only when there is a system of individual ownership and where field boundaries remain stable. In subsistence farming, however, the boundaries of a field often change completely from one year to the next.
The total area under cultivation in a given year can be measured on an aerial photograph. This is the simplest method, but an expensive one if the photos have to be taken especially for the monitoring unit. Sometimes existing photos can be used even if they were taken for other purposes. A scale of 1/10,000 will give satisfactory results.

21.3.2 Measuring on the ground

Two people are needed to measure a field: the enumerator and the farmer or two enumerators. A field will not usually have straight boundaries but its perimeter can be approximated by a fairly simple geometrical figure with straight sides. It is that simple figure which the enumerator will measure, taking the following steps:

- The enumerator makes a rough drawing of the field he is going to measure. This figure has only straight lines (Figure 21.3);

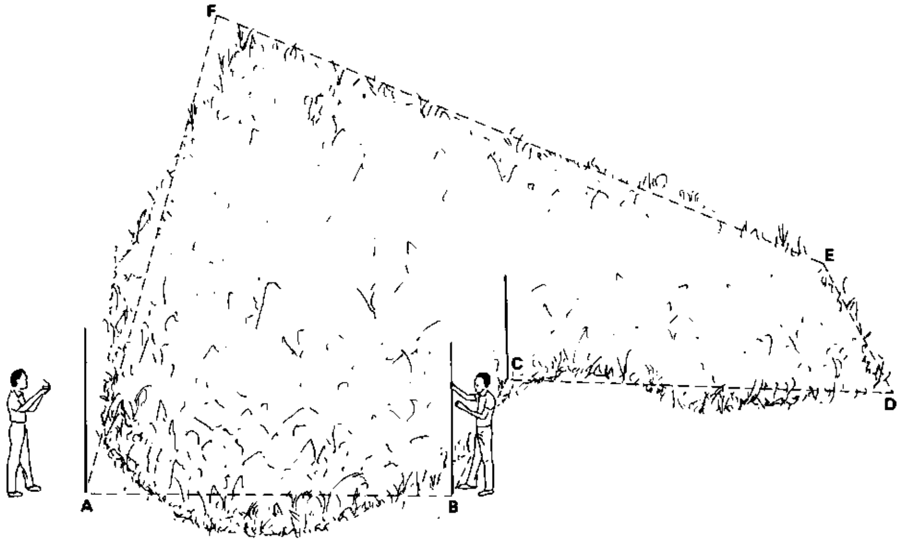


Figure 21.3. Field approximated by simple geometrical figure

- One enumerator goes to a corner of that figure (A) and the second one to the next corner (B). With a compass, the first enumerator measures the direction of the line between A and B, using the second enumerator as a reference;
- Together both enumerators measure the distance between A and B going in a straight line, not measuring the field boundary but the length of one side of the figure approximating the field;

- The first enumerator then stands at B and the second enumerator goes to the next corner (C) and the procedure is repeated for each side;
- The data should be recorded immediately on a survey form. The area can be calculated at the office soon afterwards so that wrong results can be re-measured without delay.

There are other systems of measuring fields but they are more complicated or require more fragile equipment. They all share the same disadvantage of being time-consuming if the field is not rectangular.

21.3.3 Relation between area and plants or seed

For transplanted crops, the area cultivated can be estimated if one knows the total number of plants used and their density. This can be tested on a few fields to see if a relationship seems to exist, and then used for the other fields.

A relationship between the quantity of seeds sown and area cultivated is more difficult to establish and is not very reliable anyway because it is hard to find out the quantity of seed really used.

21.3.4 Combining information

A combination of data from the sample plot and the total production figures can give a reliable estimate of the area cultivated. The sample harvest gives the production per unit of area which, combined with the total production, makes it possible to calculate the area cultivated.

21.3.5 Sources of error

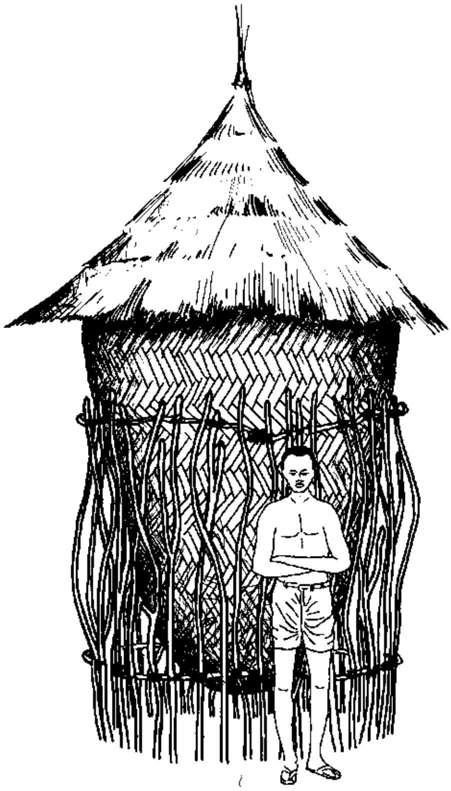
Measurements have to be checked whenever possible because of the high risk of error. The usual sources are:

- The boundaries of a field are not precise because it is not always clear where the field really starts;
- The lines measured are straight lines; the enumerator does not follow the contour of the field, which is never straight;

- Errors in reading the instruments and noting the results. This can be reduced by thoroughly training the enumerator and teaching him to write down the information systematically. The enumerator should make the drawing of the field on a sheet of paper and note the measurements and the other data on it. It is advisable to use a combined form, to be stored in a binder, for yield estimates and the calculation of the area. All the data on one crop can be put on one side of the form, while the space on the back can be used to draw the shape of the field. Just by pacing the distances of the sides of some fields makes it easy to verify whether the data are reasonable. If not, they can be measured again. A second check can be made by sending a carbon copy of the drawings to the office, where the areas are calculated. If measurements are found to be incorrect, the enumerator can be requested to do them again;
- After the harvest, an extra check can be done on those fields which show unusual production figures (very high or very low). By re-measuring the field or re-evaluating the harvest (granary method) it may be possible to eliminate some unusual data;
- If the field is cultivated in regularly spaced rows, measurements of the field can be checked by counting the rows. The length of one side, divided by the spacing between the rows, should be about equal to the number of rows on that side.

21.4 Storage of production

Farmers store much of their food crops on the farm. The storage facilities they have available influence the quantity and quality of food available for consumption, and affect the quality of the grain used for seed. Adequate storage makes it possible to sell grain when prices are most favourable.



Granary

In a project aiming at increasing food production and improving the lives of the household members, it may be useful to find out how the farmers store their crops and possibly to introduce improved storage methods. This can be a fairly sensitive matter because, traditionally, strict regulations govern access to the granaries. The attitude of the farmer towards a discussion of his stored grain can also be sensitive, depending upon the time of year; he may be reluctant to admit how little he has left during the hungry period. Data on storage should therefore be gathered by the village enumerator who knows the household situation. The survey has to be designed in cooperation with an agronomist. It should be done only if the existing storage is not effective, to identify which type of improved granary or protection measures might be introduced.

22 SURVEYS ON LIVESTOCK

Poultry and livestock are ubiquitous in Africa. Every farmer has at least some poultry and usually some goats or even cows or oxen. In some areas, farming households keep only small livestock (poultry, sheep, goats) - the cattle being reared by pastoral people. The small livestock are usually left to forage for food during the day, perhaps receiving some husks and cereals in the evening.

In traditional agriculture, poultry and livestock serve three purposes:

- As a means of savings: In a good year, surplus money is invested in livestock; in a bad year, when food shortages occur, some livestock can be sold to purchase food;
- As a means of agricultural production: Cattle can be used for animal traction, as suppliers of manure, etc.;
- As objects of trade: Although systematic trading of livestock is not a common activity among sedentary farmers, it can be of importance to some. This is especially true of poultry, which is easy to sell at the local market. (The trading of cattle by pastoral people lies outside the scope of this book and will not be discussed.) Farmers who systematically trade in livestock are more likely to take better care of their animals and to feed them regularly.

Data on livestock will be needed by the project as part of the background information required to describe the farming system and also because of the readily available source of cash that livestock represents to the farmers. If one of the project components is to improve

livestock-keeping by immunization programs and better care of the animals (which, incidentally, can produce quite spectacular and rapid results), detailed long-term surveys will be required.
Depending on the reason why data on livestock are being gathered, there will be three categories of surveys: inventory, productivity, and income. Before describing these surveys, some basic problems inherent in livestock surveys will be reviewed.

22.1 Basic problems in livestock surveys

- In many countries there is a head tax on cattle. The farmers are therefore likely to be suspicious of questions about livestock. These suspicions can be allayed somewhat by detailed, repeated explanations of the purpose of the survey and constant reassurances that the project has no connection with the government tax office;
- Keeping a herd is a means of storing wealth, so farmers are not going to give away much detail on this matter to a stranger. They will probably be reluctant to mention the real number of animals not present in the compound. It might take a long time before the enumerator has sufficiently gained the confidence of the farmers to obtain correct data. A survey on cattle should be designed with the understanding that for a long preliminary period no valid data will be obtained. It might help to avoid asking any direct questions about money, talking only about the animals.

22.2 Inventory of livestock

An inventory survey gathers data about the number and types of animals owned, if necessary distinguishing between male and female and between adult and young, noting who takes care of them and how, and for what purpose they are kept. For a detailed long-term livestock survey, the inventory data can be obtained during the first reconnaissance phase of that survey. If the inventory is merely to provide background information about the farming system, the survey can be a simple single-visit one; no further data on livestock need be gathered, except to update the

inventory each year to monitor changes.
An inventory of poultry need only be done roughly, because it is no use counting every chicken when there is a high mortality rate, with the poultry population varying greatly from one day to the next.
In making a livestock inventory, it is best to use generally accepted classifications for young and adult animals. Those recommended by FAO are presented in Table 22.1.

Table 22.1. Standard age classes in livestock inventory (after FAO)

Type of animal	Age limit young/adult in years
Buffalo	3
Camel	4
Cattle	2
Goat	1
Horse	3
Pig	0.5
Sheep	1

22.3 Productivity of the herd

If a project is to introduce improvements in livestock raising, the productivity of the herd (or the productivity of that part of the herd on which improvements will be focused) has to be evaluated. A survey on herd productivity is best implemented in stages.

22.3.1 Surveys on productivity

The first stage of a productivity survey is a reconnaissance survey to obtain an inventory of the animals and some basic information. This survey can be done on the entire population or on a large sample, depending on the work capacity of the unit. The results of the reconnaissance survey will provide a starting point for the next stage, showing the best way to select further samples and indicating the points on which the project should focus its extension programs.

A possible form for the reconnaissance survey is shown in Table 22.2. As can be seen, the questions are limited to basic information.

In the second stage of the productivity survey, the activities in livestock are monitored. The sample selected for this survey can be a random sample of the population or of those farmers who are more actively engaged in livestock.
The farmers selected should be interviewed at regular intervals. The frequency of the interviews depends on the type and validity of the data to be obtained; once a week or once a fortnight is often enough. During the interviews, information should be obtained on all events within the herd (purchases, sales, births, deaths, exchanges, etc.) and on what the farmers are doing to keep their herd in good shape (veterinary care, feed, etc.). At this stage it might be advisable not to talk yet about the money involved. The data obtained can be checked partly by an annual inventory of the herd of the sample farmers.
The results of the productivity survey can be used to make rough estimates of how much money is involved in raising livestock, on the basis of prices of animals at local markets.

22.4 Income from livestock

Once the farmers have gained enough trust in the enumerator, a survey can be designed to include the values of all the transactions. Values can be in cash or in kind, as when a cow is exchanged for other animals or given to someone else to be raised in exchange for other services.

A survey on the income from livestock will usually be part of a survey on the household's total income. The data needed cover purchases, sales, losses, and feeding expenses (in cash and kind). The income from each type of animal can be analyzed separately. Fluctuations in the size of the herd are not always relevant, except as a check on the accuracy of

Table 22.2. Form for a reconnaissance livestock survey

Name: .... Date of inquiry: ....

Location: ....

Inventory of livestock owned

Type of animal	Number at the farm				Number not at the farm			
	Male		Female		Male		Female	
	Young	Adult	Young	Adult	Young	Adult	Young	Adult
Oxen								
Cows								
Sheep								
Goats								
Donkeys								
Poultry								

Members of the household active in livestock

Name	Sex	Role in livestock

Did you prepare a fodder stock last year? Yes/No

If yes, with what? ....

At what period did your stock run out? ....

Which animals did you feed from it? ....

Did these animals get any other food? Yes/No

If yes: What was given? ....

To which animals? ....

During which period of the year? ....

Was any veterinary care given to the animals in the last year? Yes/No

If yes, to which animals? ....

the sales record.
The questions can be presented in a table because the list of possible animals and what can be done with them is limited and is known beforehand. Various tables are possible, but one that includes space for preliminary calculations of the results is especially practical. An example of such a table is presented in Table 22.3.

Example of a survey form for income from livestock

Table 22.3 actually consists of three tables: one for data on the acquisition of livestock, one for data on livestock or livestock products that leave the farm, and one for expenses incurred for the animals. Draught animals are treated separately because this information is needed for the analysis of income from crop cultivation. The form is designed for easy tabulation (precoded columns) even on a computer. During interviews with the farmers, the enumerator asks the appropriate questions and records the answers in the tables.

Table 22.3 shows that since the last interview, the farmer has not acquired any new livestock but sold one goat for 3,000 CFA at the local market. One of his cows died from an infection, so the meat was bad and could not be sold. The farmer spent 2,400 CFA on veterinary care for the sick cow. Before ploughing, he gave one 'tine' of sorghum from his stock to his draught animals.
The validity of the data can be checked by comparing the survey prices with market prices. If chicken are usually sold for 300 CFA, then an entry 'Chicken 3,000 CFA' should be questioned. Probably the enumerator put one zero too many on the price, or it was not a chicken that was sold but a sheep, or the chicken was sold to a particularly gullible expatriate. In any case the odd price has to be proven right before it can be used in analysis.

Table 22.3. Table for survey on income from livestock

ACQUISITION OF LIVESTOCK

Date	Oxen (1)		Cow (2)		Sheep (3)		Goats (4)		Poultry (5)		Others (6)		Source					Observations
interview	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	Market (1)	Birth (2)	Gift (3)	Exchange (4)	Others (5)	
12/5/81																		Nothing
Total																		Grand Total CFA

LIVESTOCK OR LIVESTOCK PRODUCTS THAT LEFT THE FARM

Date	Oxen (1)		Cows (2)		Sheep (3)		Goats (4)		Poultry (5)		Others (6)		Destination						Observations
interview	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	No	Price (CFA)	Market (1)	Gift (2)	Exchange (3)	Loss (4)	Consumed (5)	Others (6)	
12/5/81							1	3000					x						
12/5/81			1	?												x			Died from infection
Total																			Grand Total CFA

Table 22.3. (contd.)

EXPENSES FOR THE ANIMALS (food, veterinary care etc.)

Date interview	Nature of expense	Quantity	Source			Value if possible	Destination of expenses						Observations
			Market (1)	Stock (2)	Other (3)		Draught Animals (1)	Cows (2)	Sheep (3)	Goats (4)	Poultry (5)	Others (6)	
12/5/81	vet.care	–			x	2400		x					For cow that died
12/5/81	sorghum	1 tine		x			x						
Total													Grand total CFA

Observations:

23 HOUSEHOLD WELL-BEING

When the monitoring unit has completed its tabulations of agricultural production and income, it can describe the farming conditions and the results being achieved by the population in the project. It cannot yet, however, answer the basic question: Has the project improved the lives of the people? The purpose of any agricultural development project is to increase the well-being of the project population. Specific objectives such as increasing food production or creating new income opportunities are all intended to contribute to improving the lives of the people.

This chapter will review a few of the key points in estimating the well-being of the households by measuring their food sufficiency, their nutritional status, and their purchasing power.

23.1 Food sufficiency

Data on farm production are more meaningful if they are compared with the quantity of food required to feed the farming household. One ton of sorghum sounds like a lot of food, but for a household of eight people, it only amounts to 340 grams per person per day - with nothing set aside for seed, or to sell, or even to compensate for storage losses. The unit's reports on agricultural production will therefore provide better information to the users if they include data on food requirements and consumption. Such data can be used for two purposes:

- To estimate food sufficiency: The net production of each household can be compared with its standard food requirement, from which the percentage of food-sufficient households can be calculated. This can be done by taking either their food crops only or also their cash crops;

- To estimate nutritional status: The actual food consumption of the household can be estimated and compared with standard food requirements and food production.

Before these two points are discussed, it will first be explained what standard food requirements are.

23.1.1 Standard food requirements

When data on the agricultural production of a farm are being analyzed, a basic question is: Is the production sufficient to feed the household - either through direct consumption or through the sale of its cash crops to buy additional food?

A traditional kitchen

To answer that question, one first has to determine how much food the family needs to be adequately fed. This is no easy task, because what is sufficient to keep a person from going hungry does not necessarily provide that person with the balanced diet he or she needs to remain in good health.

The amount of food needed by a person depends upon the age, sex, and weight of the person, the type of work he or she does, and the climate in which he or she lives. This amount is usually expressed in calories or joules, units used to measure the energy value of food. FAO and the World Bank have set standard caloric needs per person per day for all countries; they also regularly publish data on average real consumption. For selected countries of West Africa, these data are presented in Table 23.1.

Table 23.1. Food consumption intakes and standards for selected West African countries

Country	Observed food intake in calories/ person/day (1977)	Standard requirement in calories/ person/day
Mali	2,117	2,350
Chad	1,762	2,380
Upper Volta	1,875	2,375
Niger	2,139	2,350
Benin	2,249	2,295
Sudan	2,184	2,350
Togo	2,069	2,300
Senegal	2,261	2,380
Ghana	1,983	2,305
Nigeria	1,951	2,350
Ivory Coast	2,517	2,400

From: World Development Report 1980, New York, Oxford University Press for the World Bank, Washington, D.C., August 1981 and FAO Production Yearbook 1980

The data in Table 23.1 are national averages, so they are only a rough approximation for an individual in a specific region. Still, they are sufficient to allow an estimate of the food requirements of a household (in calories). This is found by multiplying the standard requirement per person per day by the number of people in the household. This can then be compared with the household's real food intake. Conversely, the food intake figures for the country can be used to calculate the food requirements of a household if its members are to be fed at the average level for the country. This is done by multiplying the observed average food intake per person per day by the number of people in the household.

Either standard requirements or observed intake can be used in the analysis, although it should always be stipulated which one is being used. Both require that the number of people in the household be known.

The simplest way to count the people in a household is to regard each individual who is residing permanently in the household as one consumption unit. This applies also for children, to whom the same mean caloric need per day should be attributed. Small children may need less food than adults, but people who work hard physically and women who are nursing a baby need more than the average adult, which tends to even things out. FAO takes these facts into account in its standard requirements.

Caloric value of food

Both standard requirements and observed food intake are expressed in calories, so they are not directly comparable with the farm production, which is expressed either in kilograms for each crop, or in monetary value. Weights can be converted to caloric content, however, as is shown in Table 23.2, which lists the caloric content per 100 grams of some of the major food crops in the tropics.

Table 23.2. Food composition table for some common foods[1]

Type of food	Calories *Per 100 grams*
Wheat flour (medium extraction)	350
Rice (milled white)	360
Maize (grain or wholemeal)	356
Sorghum	343
Potatoes (Solanum)	70
Sweet potatoes	97
Cassava (fresh)	109
Plantains	75
Dry beans and peas	345
Groundnuts (shelled)	546
Vegetable oils (pure)	884
Sugar (refined)	387
Beef carcasses (medium)	225
Chickens dressed (not drawn)	122
Hen eggs (in shell)	144
Dried fish	223
Milk (whole cow's)	68
Milk (whole dried)	492
Skim milk (dried)	360

Source: After C. Chatfield. *Food composition tables for international use.* FAO Nutritional Studies No. 3, Washington, D.C. 1949

[1] Based on retail weight of the foods as purchased

This table can be used in two ways:

- To calculate the caloric value of a given quantity of food;
- To calculate the quantity of food necessary to provide a required number of calories.

In the rural areas of many tropical countries, one type of food (e.g. rice, sorghum, cassava) provides most of the caloric intake of the households, so calculations can be simplified without too much error. The cost of the staple food can also be used to estimate the monetary value of food required by the households.

Example of calculating the caloric value of food

In a country where the staple food is sorghum, the heads of the households usually take out of the granary the same amount of sorghum each day for cooking. In a household of six people, the head may take out, say, 3.5 kg of sorghum each day. The number of calories available to the family daily can then be calculated: $\frac{343 \times 3{,}500}{100} = 12{,}000$ calories (rounded off). The quantity of sorghum is converted from kilograms to grams because the caloric contents in Table 23.2 are given per 100 grams of food. Thus the number of calories available per person per day is: $\frac{12{,}000}{6} = 2{,}000$. This figure can be compared with the observed and standard requirements for that country.

Example of calculating the quantity of food required

Suppose rice is the staple food in a country where the observed food intake per person per day is 1,980 calories. Since rice provides 360 calories per 100 grams, each person needs $\frac{1{,}980}{360} = 5.5$, or 550 grams per day. A household of six persons needs $550 \times 6 = 3{,}300$ grams per day (3.3 kg). It needs to produce at least $3.3 \times 365 = 1{,}200$ kg (rounded off) of rice each year.

Cash value of food requirements

The examples above assumed that the households rely only on their food crop to feed themselves. This is not always true. Households will also sell some crops to buy other food. It is then necessary to compare the net income of the farm production with the monetary value of the food required by the household.

Example of comparing net farm income with food requirements

A farm's net income from crop production is 127,000 CFA. Six people have to be fed from this farm. The staple food is sorghum, which sells for 58 CFA/kg and provides 343 calories per 100 gram or 3,430 calories per kg. To meet the standard daily food requirement of 2,350 calories, each person needs $\frac{2,350}{3,430}$ = 0.685 kg of sorghum per day. A household of six persons will need 0.685 × 6 = 4.1 kg per day (rounded off). At 58 CFA/kg, the household needs 58 × 4.1 = 238 CFA per day for food, or 238 × 365 = 86,870 CFA a year. The net income of 127,000 is sufficient to meet these needs and still leaves some 40,000 CFA for other expenses.

Naturally, such a calculation is only a rough estimate. The family eats food other than sorghum, and it also has sources of income other than crop production. Even so, this calculation is a useful indicator of whether a farm has produced enough to feed its workers and their dependants.

Sometimes a project may set standard food requirements, which may be, say, an annual diet per person of 240 kg of sorghum and 30 kg of cowpeas. The total value of the food required by a household can then be calculated directly. If sorghum costs 58 CFA per kg and cowpeas cost 79 CFA per kg, it will cost 240 × 58 + 30 × 79 = 16,290 CFA to feed one person, or 97,740 CFA for a household of six.

Once the farm production and the food needs of each household have been compared, it is easy to prepare a frequency table showing the percentage of food-sufficient households. This can be helpful to the project managers in identifying the location and extent of insufficient food production. A frequency table, updated over the years, can illustrate the success (or otherwise) of a project.

23.1.2 Nutritional status

A detailed study of the nutritional status of the household would be difficult, if not impossible, for a monitoring unit to conduct. It would mean gathering precise data on the quantity and type of food consumed by each member of the household, and would involve weighing the food and analyzing its content. Rough estimates, however, can be made by estimating the food intake per capita. If anything more than rough estimates are required, it is best to obtain the help of a nutrition expert, who will suggest appropriate methodologies.

A basic constraint in gathering food intake data is the sex of the enumerator. Whereas the head of a household, who is most likely a man, knows the quantity of staple foods needed to feed the household, it is the women who cook and know how much food is actually consumed. In many societies, it is impossible for a male enumerator to obtain information directly from the women, so if the unit does not employ female enumerators, this can constitute a problem. A possible solution may be for the unit director to arrange for female extension agents to gather this type of data.

The method described in this section assumes that the work is done by female enumerators, who will conduct a survey on a sample of the project population. At regular intervals throughout the year, they will interview the sample households, asking what type of food was eaten on the day before the interview, in what quantity, and how many people were there to share it. The quantities will be approximate, in local units of measurement.

In areas where the staple food is easy to measure, it may be helpful to give each sample household a set of tin dishes of different colours to use when taking the grain for cooking. These dishes can be filled and weighed with different food crops beforehand, so that a table of average weights can be compiled.

The enumerator should insist that all types of foods be mentioned, not just the staple foods or the food that was cooked. The women may not think of mentioning fruits, leaves, or roots gathered on the way to the field, or food eaten outside of the home (e.g. at the market).

The survey should take place several times throughout the year because food intake can vary with the seasons. Four times a year should be sufficient. Care should be taken to avoid festival days and periods of fasting, such as Ramadan in Moslem countries.

Conducted too frequently, say once a week, a food intake survey can annoy the women, especially if there are times when little food is available. Done only four times a year, there is little risk of causing any ill feeling.

Another advantage of only a few interviews a year is that the information can be entered on one survey form, allowing the average consumption to be calculated on the form itself. The average total intake per person per year for all the households in the sample can then be compared with the average farm production per person and the standard food requirements for the area.

During the third or fourth interview, when the women are used to being questioned, the survey could also include a few open-ended questions to identify:

- The types of food most often eaten;
- Where most of the staple foods are obtained (own production or bought);
- What type of food is considered very desirable but is never, or only rarely, eaten, and why (cost, availability).

23.2 Household purchasing power

Food is not the only basic need of a household. Every household needs money to cover such expenses as housing material, clothes, cooking implements, taxes, and so on. Most households earn money from several sources: selling their crops, raising livestock, or through some activity such as handicrafts or petty trade.

To find out whether the households in a project are able to meet their monetary needs, no detailed data on their total income and total expenses are needed. This would be an impossible exercise anyway, but rough estimates of the total household income and total household expenses in cash and kind will usually suffice.

A note of warning is in order because data on income and personal expenses are difficult to obtain, and for three reasons:

- Income and personal expenses are touchy subjects. Many people are reluctant to talk about their money. There are often strong social incentives to keep quiet about one's wealth or lack of it. Calling the attention of the tax collector, jealous neighbours, or hungry relatives to one's savings is unwise. Calling the attention of disparaging neighbours and relatives to one's lack of success is humiliating. Farmers who are very cooperative in a survey on their farming practices may baulk when asked about income and household expenses. Some survey specialists have had more success in obtaining data on expenses than data on income;

- Details of income and expenses are difficult to remember. Few people can remember precisely everything they earned or spent for very long. A detailed study of household expenses would require frequent interviews and would soon bore both the enumerator and the farmer;
- More than one person is involved. In many households, various persons have some source of income and these persons all incur expenses for the household. The head of the household may be aware of the major transactions conducted by these persons, but not of all the minor ones. To have a full set of data, the enumerator would have to conduct regular interviews with all persons concerned, including the women. Some of these persons may be reluctant to divulge their income in front of the head of household.

23.2.1 Estimating the total income of a household

Estimating income (in cash and kind) from cultivation and livestock activities has been discussed in earlier chapters and will not be reviewed here. Some suggestions for obtaining data on the other sources of income are given below.

Petty trade

Many farming households regularly engage in petty trade, i.e. selling manufactured goods on a small scale. While the profit margin is likely to be small, trading constitutes a steady source of income and can be quite important in certain households.

Data on petty trade are difficult to obtain on a large sample, as the trader should be interviewed frequently (probably twice a week). The survey can best be designed in two phases. In the first phase, which could last for one year or one crop season, a few questions on trade could be asked on a sample representative of the population. The answers should reveal what type of trading is done, on what scale, and which groups are particularly active in it. The second phase could be a more detailed survey done on a sample of active traders.

If some traders can write, even if only a little (or if one of their children can write), it may be possible to prepare a notebook in which the trader (or the child) can record the transactions. This is feasible only in the second (detailed) phase of the study. Traders are likely to be quite cooperative in this because most of them are interested in finding out how much they actually earn.
Such a survey, however, is difficult to do and is time-consuming at first, more so than a normal survey. The enumerator will have to spend a lot of time with each trader, showing him or her how to use the notebook in the course of a business day. He will need to come back frequently until the trader has properly understood the system.

Handicrafts and processing of farm produce

In subsistence farming, it is a rare household that does not engage in some form of handicraft, either for personal use or for sale. As well, farm produce is often processed and sold, mostly by the women. Data on these activities are needed not only to estimate the total income of the household, but also to identify areas in which the income of the household could be raised through training or credit programs. These data will also form the basis for any future development of small-scale agro-industries.

Earning income from handicrafts

Methodologies similar to those described for petty trade can be applied to gather data, again in a two-phase study.

Paid off-farm work

Paid work outside the household farm sometimes contributes to the household income. Household members may sometimes work for another farmer, a local shop, a construction site etc., or household members who have gone to work in town may send money to their people back home. In an area of predominantly subsistence agriculture, both these sources can mean considerable additions to the household income. They therefore constitute important items of information, but are difficult to obtain. The problem is not one of recall - people remember well how much money they received - but one of willingness to tell.

Questions about paid work could be included in a set of questionnaires on petty trade, income from handicrafts, and the processing of farm produce. The questions should cover the amount earned and type of work done. This will broadly identify the categories of work available in the area. Informal inquiries can reveal their going rate.

Data on income should preferably be noted on the questionnaire only after the income has actually been received, not while the work is still being done.

23.2.2 Real expenses

Many surveys on household budgets are designed to include every household expense. Such surveys are troublesome to conduct and the data they produce are only of dubious quality. Besides, project managers will seldom need full details on household budgets, but merely a rough indication of whether the households can meet their basic expenses and what their financial capabilities are. A few informal inquiries about regular expenses - asking people how often they renew their household goods and clothes, what taxes they pay, and so on - will usually suffice. The going price for frequently bought objects can be checked in the shops.

23.2.3 Minimum cost of living

Most monitoring units will be required to find out what the minimum needs of a household are and how much these minimum needs cost. This information can be gathered over the months by informal interviews with male and female members of the households and also with male and female members of the project staff. Female project staff (enumerators, health workers, extension agents), in particular, are aware of the extent and type of goods owned by the households and are familiar with prices at the local shops and markets.
When building up a 'typical' household budget, the household size has to be taken into account because a distinction will be made between individual goods, which have to be duplicated for each member of the household, and household equipment.

> For example: If it is customary for a man to buy a dress for his wife after harvest, a man with three wives will buy three dresses. On the other hand, while a household of ten needs more cooking pots than a household of two, it does not need five times as many.

It is impossible to give a standard list of needs because these obviously vary with location and tradition. The following list serves merely as a reminder of the usual categories of household expenditures; it does not include expenses for agricultural production:

- Food and drink: including cost of festivities (religious days, weddings, funerals);
- Housing and furniture, maintenance and repair of buildings;
- Cost of running the household: water, wood, fuel, cooking implements, food storage facilities;
- Transportation and travel;
- Clothing and personal objects;
- Medical care;
- School expenses;
- Taxes;
- Social obligations (in cash and kind);
- Cost of credit for household goods and food.

Once a standard list of expenses has been compiled, it should be discussed with farming households and people familiar with the area, and modified if necessary. It can then be used to draw up a 'minimum cost of living' for households of diverse sizes. When income figures are presented in reports, they will be more readily understood if accompanied by figures on the minimum cost of living.

24 DATA PRESENTATION IN FREQUENCY TABLES

Presenting data in long lists would confront the reader with a confusing mass of information which he would find difficult to digest. A far better approach is to present the data in tables. These have the advantage of allowing a great deal of information to be presented in a way that is easy for the reader to understand. He can see at a glance in what range most of the data fall and whether there are many extreme values.

Tables are often copied from the report they are in and used independently of the report. They should therefore be compiled in such a way that they can be understood - on their own - by anyone. The heading of a table should state clearly what information the table contains and, if relevant, should mention the source of the information. The headings of the columns should speak for themselves or be explained in a footnote to the table.

The tables most commonly used to present data on farm surveys are frequency tables. These show the distribution of the frequency of occurrence of a given characteristic of the population, grouped into categories or classes.

24.1 Absolute frequency tables

The simplest form of frequency table is the absolute frequency table. It records the total number of entries in each category or class and presents them in columns in the table.

For example: If the farmers are asked whether they use animal traction or not, their answers will be 'Yes' or 'No'. The frequency table showing this information will have two categories, one for 'Yes' and one for 'No'. The absolute frequency of each answer (i.e. the total number) will be recorded in the appropriate column of the table.

An example of a more complex frequency table - but one still easy to read - is that showing the number of children that each woman in a sample has given birth to. Such a table would need categories from 0 to perhaps 15, with the total number of women in each category represented in the appropriate column of the table.

If the number of categories becomes too large, the table defeats its purpose of being easy to read. This problem can be overcome by grouping the entries into classes and recording the frequency of entries in each class. This reduces the size of the table to reasonable proportions.

The ideal classification provides enough information (not too few classes) without becoming too large (not too many classes). In choosing a classification, one should keep the following in mind:

- There should be at least five classes;
- The classes should be of the same size (although sometimes this principle can be violated, as will be explained later);
- The frequencies should increase from the first class onwards, should reach a peak, and should then decrease;
- The class boundaries should be round numbers.

In choosing classes for numerical data, there are no exact rules for the correct choice of classes, although they are usually chosen in the following way:

- Deduct the lowest value from the highest value; this provides the range of values;
- Divide this range by the number of classes desired (at least five) and round off the value obtained.

If the number of data to be classified is small, the number of classes has to be kept small; if the data are many, more classes can be used.

Because the class boundaries are rounded off to keep them simple, the total range of classes may exceed the range of the highest minus the lowest value. For this reason, the first class may start below the lowest value and the last class may end above the highest value. The difference between the range of values and the range of classes should be divided roughly between the first and last class.
After the classes have been chosen, the frequency of entries in each class is calculated. The frequency of all the classes should be added up. The sum should be equal to the number of entries. The distribution should have a regular shape, increasing from the first class, having only one peak, and then decreasing. If the distribution is erratic, a change in class boundaries or a reduction in the number of classes can perhaps regulate this. If that does not help, it may be that the data are not from a homogeneous population.

For example: Suppose the weight of cows is surveyed and tabulated (Table 24.1). This table shows two peaks in the frequency distribution: the first peak occurs in the first class, the second in the fifth class. A probable explanation for these two peaks is that the original population included both cows and calves, so a stratified sample should have been taken.

Table 24.1. Weight of a sample of 36 cows

Weight in kg (classes)	No. of cows (frequencies)
0 - <* 200	10
200 - < 400	6
400 - < 600	3
600 - < 800	3
800 - < 1,000	7
1,000 - < 1,200	5
1,200 and more	2
Total	36

* The sign '- <' means up to but not including the value that follows the sign. The first class thus contains values from 0 to 199.999....

A solution to the problem of data that are not homogeneous is post-stratification, meaning that the data that rightly belong in different strata are split up, as is shown in Tables 24.2 and 24.3. It is not always possible to post-stratify, however, because the information needed to do so may not be available.

Table 24.2. Weight of a sample of 20 cows

Weight in kg (classes)	No. of cows (frequencies)
200 - < 400	1
400 - < 600	2
600 - < 800	3
800 - < 1,000	7
1,000 - < 1,200	5
1,200 and more	2
Total	20

Table 24.3. Weight of a sample of 16 calves

Weight in kg (classes)	No of calves (frequencies)
0 - < 200	10
200 - < 400	5
400 - < 600	1
Total	16

The process of classifying data and presenting them in a frequency table will be illustrated in the following example.

Example of an absolute frequency table

The sorghum yields obtained by a sample of farmers in six villages in 1979 were listed on the original recapitulative form as follows:

Table 24.4. Sorghum yields of 72 farmers in Villages 1 to 6 in kg/ha in 1979

Village					
1	2	3	4	5	6
1,160	930	1,000	900	810	1,460
940	1,140	870	870	260	940
980	470	920	750	1,480	1,580
410	750	630	920	1,210	690
680	1,340	1,040	640	650	1,470
880	620	1,150	680	510	1,500
950	790	1,250	750	680	1,610
250	990	430	290	440	650
730	940	550	500	420	1,460
1,520	1,530	930	1,250	640	770
820	1,060	560	1,050	260	1,070
990	750	830	1,020	720	1,090

The range of values is 1,610 - 250 = 1,360. Since there are only 72 values, the number of classes should not be too large. The numbers are rounded off (units of 10) so a classification into 7 classes with widths of 200 kg/ha is a good possibility. This gives easy class boundaries and results in the following frequency of values per class. For ease of presentation, the first class starts at 250 instead of 200.

Table 24.5. Frequency distribution of the sorghum yields of 72 farmers in Villages 1 to 6 in 1979

Sorghum yields in kg/ha (classes)	No. of farmers (frequencies)
250 - < 450	8
450 - < 650	9
650 - < 850	17
850 - < 1,050	18
1,050 - < 1,250	8
1,250 - < 1,450	3
1,450 - < 1,650	9
Total	72

There are two peaks: the first in the class 850 - < 1,050, and the second in the class 1,450 - < 1,650. If this were due to random factors, it would be eliminated by a different choice of classes. But here a classification in larger classes does not solve the problem, which shows that the samples may not have been drawn from a homogeneous population.

A solution may be to segregate the population into different homogeneous sub-populations, if possible. Simply by looking at the original list of data, one can see that in Village 6 the yields are systematically higher. Examining the observations made in the field during the agricultural season, one finds that those farmers also had a higher level of technology (better crop care etc.). It seems that this had a noticable impact on the crop yields. Now Village 6 is separated from the rest, and a frequency table of the sorghum yields of Villages 1 to 5 is constructed. The five villages have a different range, 1,530 - 250 = 1,280, and sixty values. If the class width of 200 is chosen as before, the first class should begin below the lowest value and the highest class should end a little above the highest value. For ease of presentation, the first class starts at 200 instead of 250. The frequency distribution for Villages 1 to 5 then takes the following form.

Table 24.6. Frequency distribution of the sorghum yields of 60 farmers in Villages 1 to 5 in 1979

Sorghum yields in kg/ha (classes)	No. of farmers (frequencies)
200 - < 400	4
400 - < 600	9
600 - < 800	15
800 - < 1,000	17
1,000 - < 1,200	8
1,200 - < 1,400	4
1,400 - < 1,600	3
Total	60

This result shows a better distribution, with the frequency going up from the first class, having only one peak, and going down afterwards.

A class width of 190 could have been chosen, but the 200 class is easier to read. Table 24.7 shows how awkward a class of 190 is.

Table 24.7. Frequency distribution of the sorghum yields of 60 farmers in Villages 1 to 5 in 1979

Sorghum yields in kg/ha (classes)	No. of farmers (frequencies)
220 - < 410	4
410 - < 600	9
600 - < 790	14
790 - < 980	15
980 - < 1,170	11
1,170 - < 1,360	4
1,360 - < 1,550	3
Total	60

This example also shows that the choice of class boundaries determines the frequencies in the table. The general tendency, however, remains the same, independent of the classification.

It is not always necessary for the classes to be of the same size. In the case of income distributions, for instance, when there are many small incomes and only a few large ones, a classification into equal classes would give many classes with only a few values in most of them. It is then better to present the data in classes of unequal size. To allow comparisons between the different classes, the frequency density per standard class can be presented as well. This is obtained by dividing the number of values in a class by the number of times that the standard class fits into the class considered.

Example of a frequency table with unequal classes

The agricultural income (in CFA) of 33 farmers in 1978 is presented in Table 24.8.

Table 24.8. Agricultural income (CFA) of 33 farmers (1978 data)

34,800	35,100	29,700
45,100	70,400	79,800
43,900	77,000	52,900
50,900	48,500	40,200
41,300	24,600	65,700
39,900	111,200	47,300
56,400	28,800	30,300
53,500	33,800	51,800
28,900	28,000	43,500
95,000	50,400	42,500
35,500	59,100	86,000

The range of observations is 111,200 - 24,600 = 86,600. Nine classes of 10,000 would therefore seem appropriate, with the lowest class starting at 20,000. The frequency distribution is presented in Table 24.9.

Table 24.9. Frequency distribution of the income (CFA) of 33 farmers (1978 data)

Income per farmer in CFA (classes)	No. of farmers (frequencies)
20,000 - < 30,000	5
30,000 - < 40,000	6
40,000 - < 50,000	8
50,000 - < 60,000	7
60,000 - < 70,000	1
70,000 - < 80,000	3
80,000 - < 90,000	1
90,000 - < 100,000	1
100,000 - < 110,000	-
110,000 - < 120,000	1
Total	33

The incomes above 60,000 CFA are too few to justify the six extra classes. A better classification, using a class width of 10,000 CFA as standard class, is the following:

Table 24.10. Frequency distribution of the income (CFA) of 33 farmers (1978 data)

Income per farmer in CFA (classes)	No. of farmers (frequencies)	No. of farmers per class of 10,000 CFA (frequency density)
20,000 - 30,000	5	5
30,000 - 40,000	6	6
40,000 - 50,000	8	8
50,000 - 60,000	7	7
60,000 - 80,000	4	2
80,000 - 120,000	3	0.75
Total	33	

In selecting classes with unequal class widths, three points have to be considered:

- The unequal classes must be larger than the standard class;
- The class with the highest frequency should be a standard class;
- The class width should only become wider after the maximum frequency has been reached and should not be narrowed again afterwards.

It is also possible to keep the highest class open, especially if that class is going to be large. In the example, the highest class could be denoted as 80,000 or more. This provides less information to the reader but can be useful if the higher values show a wide scatter.

Up to now, all classes have been denoted by the '- <' sign, the most commonly used notation. When classes cover only a small range of indivisible units (e.g. the number of people in households), the classes can be denoted by the '-' sign. A class of 0 - 4 means that the number 4 is included; since no value can occur between 4 and 5, the next class will be 5 - 9, and so on.

24.2 Relative frequency tables

In addition to, or in combination with, absolute frequency tables, data can be presented in other frequency tables. One of these is the relative frequency table. A relative frequency table presents the absolute data converted to percentages. This has the advantage that different numbers of absolute data can be compared with one another, although only if their class boundaries are the same.

Example of a relative frequency table

From the absolute frequency of the sorghum yields in Villages 1 to 5 (Table 24.6), the relative frequency is shown in Table 24.11.

Table 24.11. Frequency distribution of the sorghum yields of 60 farmers in Villages 1 to 5 in 1979

Sorghum yields in kg/ha (classes)	No. of farmers (absolute frequency)	Percentage of farmers (relative frequency)*
200 - < 400	4	7
400 - < 600	9	15
600 - < 800	15	25
800 - < 1,000	17	28
1,000 - < 1,200	8	13
1,200 - < 1,400	4	7
1,400 - < 1,600	3	5
Total	60	100

* rounded off

If the sample has been well chosen and can be regarded as representative of the population, the percentages in the relative frequency table can be extrapolated to apply to the population as a whole. The absolute frequency table of the population can then be compiled by multiplying the size of the population by the relative frequency of each class.

24.3 Cumulative frequency tables

Both absolute and relative frequency tables can be converted into cumulative frequency tables. These show the total number of data (or the percentage thereof) below the upper boundary of each class. Such tables are useful when one wants to find out how many of the data lie below a certain value.

For example: A project aiming at a certain minimum income for its farmers will want to know how many of the farmers have incomes below this level.

A cumulative frequency table can be prepared from either an absolute frequency table or a relative frequency table. The cumulative frequency table notes, not the classes, but the upper boundaries of each class. The cumulative frequency for each class boundary is the sum of the frequencies of all the classes below that class boundary.

Example of a cumulative frequency table

From the absolute and relative frequencies of the sorghum yields of the farmers in Villages 1 to 5 (Table 24.11), the cumulative distribution, both absolute and relative, is shown in Table 24.12. For the first class, the upper boundary is 400 and the frequency is 4. The cumulative absolute frequency of 600 is the sum of the frequencies of Classes 200 - < 400 and 400 - < 600. It is 4 + 9 = 13. The other cumulative absolute frequencies are calculated in the same way. The cumulative relative frequencies represent the cumulative absolute frequencies converted into percentages.

Table 24.12. Cumulative frequency distribution of sorghum yields of 60 farmers in Villages 1 to 5 in 1979

Sorghum yields in kg/ha (upper class boundaries)	No. of farmers (cumulative absolute frequency)	Percentage of farmers (cumulative relative frequency)
400	4	7
600	13(4+9)	22(15+7)
800	28	47
1,000	45	75
1,200	53	88
1,400	57	95
1,600	60	100

25 DATA PRESENTATION IN GRAPHS

A frequency table is not always the best way to present data in a report. A graph is sometimes a much better way of presenting data because it helps the reader to visualize the distribution of the data. A graphic presentation is less precise than a frequency table, but helps to make the report more readable and therefore more accessible to a larger number of people - an important asset in the farm monitoring process.
Some of the graphs most commonly used to present frequency distributions are discussed below.

25.1 Frequency histograms

A frequency histogram is constructed from an absolute or a relative frequency table. The figure constructed is exactly the same whether for relative or absolute frequencies; only the frequency numbers change. Like all graphs, a frequency histogram has a horizontal and a vertical axis. Along the horizontal axis, the class boundaries are marked; above them, rectangles, proportional in area to the frequency of each class, are erected.

The ideal histogram is a squarish figure, presenting the range of classes on a horizontal axis which is about the same length as the height of the rectangle presenting the highest frequency.

Frequency histograms can be constructed with classes of the same size or with classes of different size.

25.1.1 Frequency histograms with the same sized classes

Example of a frequency histogram with the same sized classes

A frequency histogram of the cotton yields of 100 farmers is to be constructed from the absolute frequency table presented in Table 25.1.

Table 25.1. Absolute frequency table of cotton yields of 100 farmers in 1979

Cotton yields in kg/ha (classes)	Number of farmers (frequencies)
0 - < 250	1
250 - < 500	5
500 - < 750	19
750 - < 1,000	35
1,000 - < 1,250	17
1,250 - < 1,500	10
1,500 - < 1,750	8
1,750 - < 2,000	3
2,000 - < 2,250	2
Total	100

The resulting histogram is shown in Figure 25.1.
The relative frequency histogram is exactly the same figure. All one has to do to obtain the relative frequency histogram is to convert the absolute frequencies into relative frequencies and mark them on the figure.

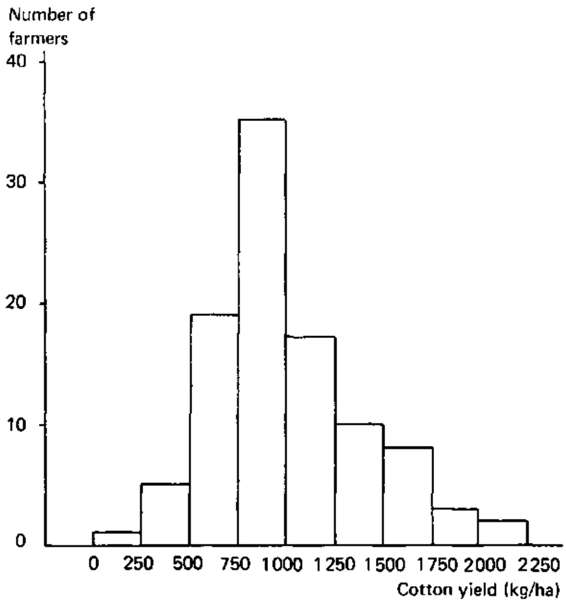
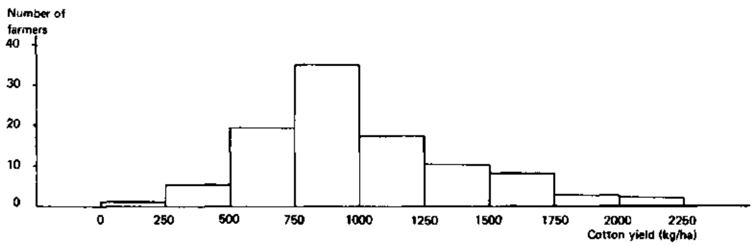


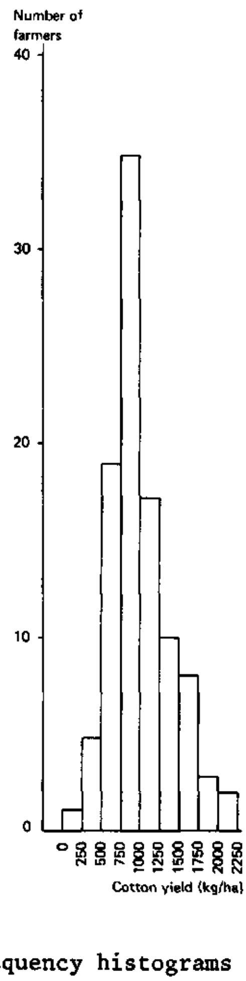
Figure 25.1. Frequency histogram of the cotton yields of 100 farmers in 1979

25.1.2 Inappropriate scales for frequency histograms

Histograms must be drawn to appropriate scales; otherwise they can easily be misinterpreted.
Figure 25.2 is a histogram presenting exactly the same data as in Figure 25.1, but drawn to an inappropriate scale. Anyone looking at Figure 25.2 and seeing its narrow class bases and the exceptionally high rectangles erected above them could easily draw the conclusion that the data are far more dispersed than they are in reality.

Another example of a histogram drawn to an inappropriate scale is Figure 25.3. It, too, presents the same data as in Figure 25.1, but anyone looking at its wide class bases and the low rectangles above them could easily conclude that the data are far less dispersed than they are in reality.





Figures 25.2 and 25.3. Frequency histograms of the cotton yields of 100 farmers in 1979 (inappropriate scales)

25.1.3 Cumulative frequency histogram

A cumulative frequency histogram is constructed from a cumulative frequency table - either absolute or relative. Here again, both tables produce the same figure.
To construct the cumulative histogram, the same process of marking the class boundaries on the horizontal axis and erecting rectangles above them is followed. Now, however, the area of the rectangles is proportional to the cumulative frequency of the upper boundary of each class.

Example of a cumulative histogram

The cumulative frequency histogram of the cotton yields of the 100 farmers is constructed from Table 25.2 and is presented in Figure 25.4.
In this graph, the cumulative frequency of 0 kg/ha (beginning of the first class) is 0. The cumulative frequency of 250 kg/ha is 1, and so on.

Table 25.2. Cumulative frequency distribution of the cotton yields of 100 farmers in 1979

Cotton yields in kg/ha (upper class boundaries)	Number of farmers with lower yield (frequencies)
250	1
500	6
750	25
1,000	60
1,250	77
1,500	87
1,750	95
2,000	98
2,250	100

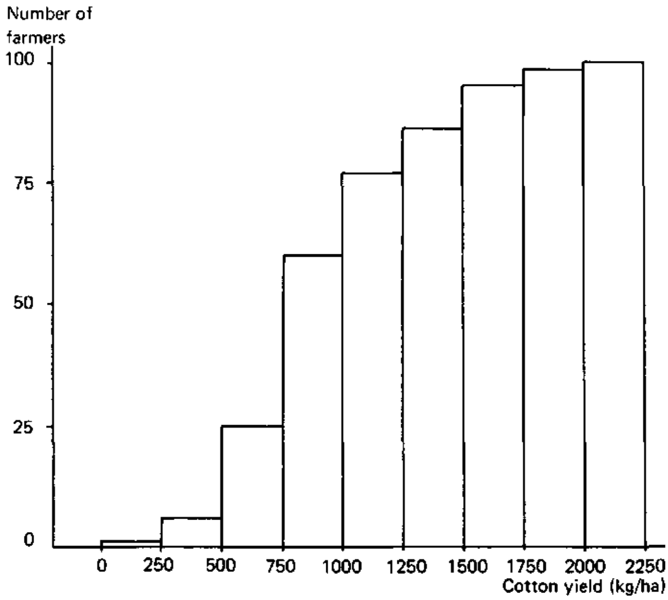


Figure 25.4. Cumulative frequency histogram of the cotton yields of 100 farmers in 1979

25.1.4 Frequency histograms with different sized classes

If the classes of the frequency distribution are of different sizes, the rectangles should be constructed in such a way that the area of the rectangle still represents the frequency. For standard classes, the height of the rectangle equals the frequency. For wider classes, the height of the rectangle is equal to the frequency density, which is the frequency of that class divided by the number of times the standard classes fits into that class.

Example of a frequency histogram with different sized classes

For this example, we shall use the data presented in Table 25.3, which shows the distribution of income from livestock of 50 farmers in 1980. A frequency histogram of this type is constructed in the same way as a normal histogram for the first three classes; for the other classes, the height of the rectangles is not indicated by the frequency but by the frequency density. The resulting histogram is shown in Figure 25.5.

Table 25.3. Frequency table of income from livestock of 50 farmers in 1980

Income in CFA (classes)	Number of farmers (frequencies)	Number of farmers per class of 2,000 CFA (frequency density)
0 - < 2,000	10	10
2,000 - < 4,000	18	18
4,000 - < 6,000	8	8
6,000 - < 10,000	6	3
10,000 - < 15,000	5	2
15,000 - < 25,000	2	0.4
25,000 and over	1	0
Total	50	

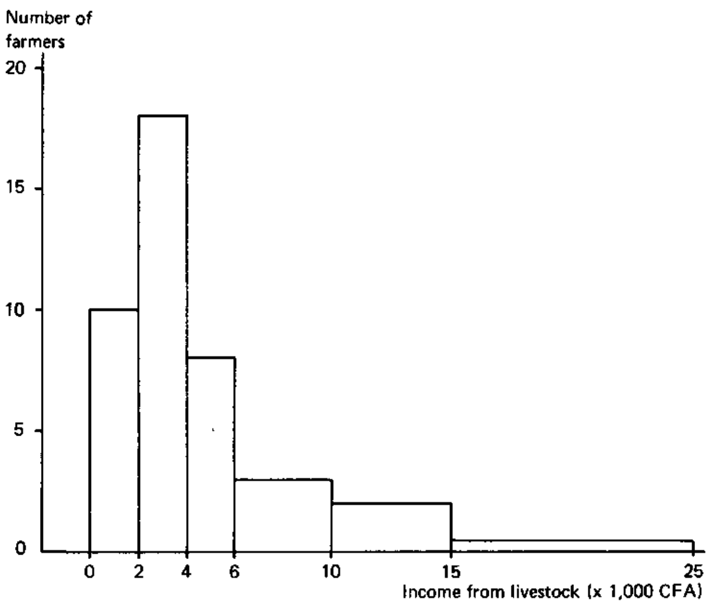


Figure 25.5. Frequency histogram of the income from livestock in CFA of 50 farmers in 1980

If the final class in the frequency table is denoted as an open class, that class can be considered to have a frequency density of zero.

25.2 Frequency polygon

Another graph that can be used to represent frequency distributions is the polygon. A polygon helps in visualizing the shape of the frequency distribution. It provides exactly the same information as a histogram, the difference being simply a matter of presentation. How to construct a polygon is explained below.

- Use the horizontal axis for the classes and the vertical axis for the frequencies;
- Identify the middle of a class and imagine a vertical line proportional to the frequency. (With unequal classes, use the frequency density instead of the frequency.) The easiest way is to place a ruler vertically against the middle of the class and follow it up;
- Mark the point found on the paper and follow the same process for each of the other classes;
- After the points have been marked for each class, connect them with straight dashed lines. The middle of the first and last class is connected with the horizontal axis at the middle of the next imaginary class, whose frequency is always zero;
- The figure should be about square (range of classes equal to the largest frequency);
- Mark clearly on each axis which variable is represented and which unit of measurement is used.

When reading a polygon, one should remember that the only meaningful points are those located at the middle of each class. The lines drawn between those points have no meaning.

Example of a frequency polygon

The frequency distribution of the cotton yields of the 100 farmers was presented in Table 25.1 and Figure 25.1. The same data can be presented in a polygon. The middles of the classes and the frequency of each class are identified as shown in Table 25.4.

Table 25.4. Middles of the classes of cotton yields of 100 farmers in 1979, and their frequencies (as presented in Table 25.1)

Cotton yields in kg/ha (middles of classes)	Number of farmers (frequencies)
125	1
375	5
625	19
875	35
1,125	17
1,375	10
1,625	8
1,875	3
2,125	2
Total	100

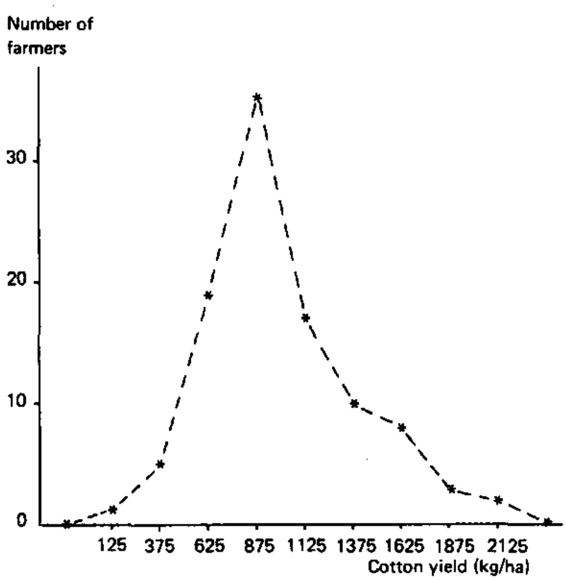


Figure 25.6. Frequency polygon of the cotton yields of 100 farmers in 1979

As shown in Figure 25.6, the class middles are marked on the horizontal axis and their frequencies are marked vertically above them. The points are connected with a straight dashed line. This line is connected with the horizontal axis at the middles of the imaginary classes below the first class (- 250 - < 0, which is - 125) and above the last class (2,250 - < 2,500, which is 2,375). The lines between the points are meaningless.

25.2.1 Cumulative frequency polygon

A cumulative frequency polygon is constructed from a cumulative frequency table. Its construction is done in much the same way as a frequency polygon except that now the upper boundaries of the classes are indicated and connected with straight dashed lines. The line starts at 0 (beginning of the first class) and ends at the point corresponding to the end of the last class. Hence the end of the polygon is not connected with the axis. The distance from a point to the axis gives the cumulative frequency of that value. Here, the line between points has a meaning because interpolation between points is possible, as it is assumed that the data are equally distributed within the classes.

Example of a cumulative frequency polygon

For the construction of a cumulative frequency polygon of the cotton yields of the 100 farmers, all the data needed are found in Table 25.2. The polygon is constructed by marking the class boundaries and their related cumulative frequencies on the graph and connecting all the points with a straight dashed line. The end of the line is *not* connected with the horizontal axis, but the beginning *is* because the first class boundary has a cumulative frequency of 0. The resulting polygon is presented in Figure 25.7.

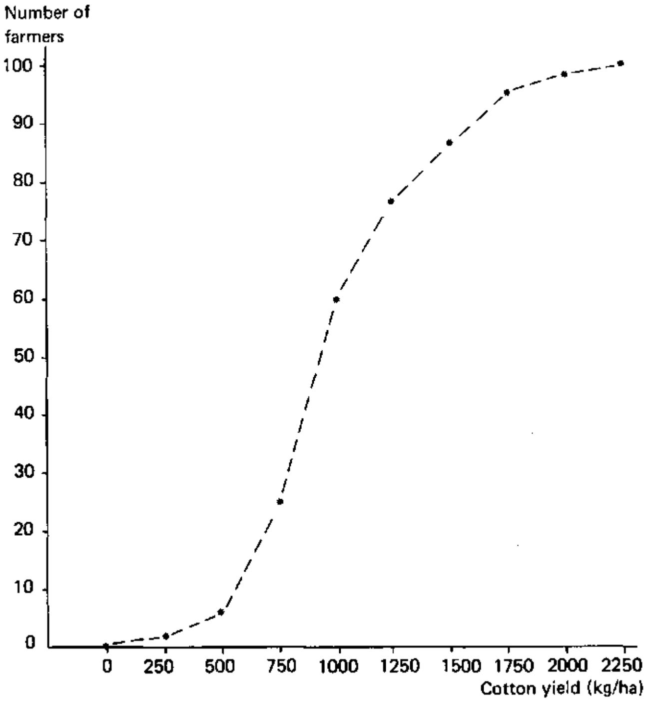


Figure 25.7. Cumulative frequency polygon of the cotton yields of 100 farmers in 1979

25.3 Needle graphs

If the topic surveyed has only a small number of possible alternatives (e.g. 'Yes' and 'No', or the number of children per woman), the frequency histogram can be replaced by a needle graph. As before, the frequency is marked on the vertical axis and the variable on the horizontal axis. Instead of rectangles, vertical lines are erected, their length representing the frequency of that value of the variable. The tops of the lines are not connected.

Example of a needle graph

A census survey was conducted in a village in 1976. One of its findings was the number of children under the age of six in each household. The figures are presented in Table 25.5.

Table 25.5. Frequency of households with a specific number of children under the age of six in a village in 1976

No. of children under six years	No. of households (frequencies)
0	5
1	12
2	16
3	20
4	44
5	48
6	76
7	80
8	40
Total	341

Although this frequency distribution could be represented in a frequency histogram, a more appropriate tool is the needle graph shown in Figure 25.8.

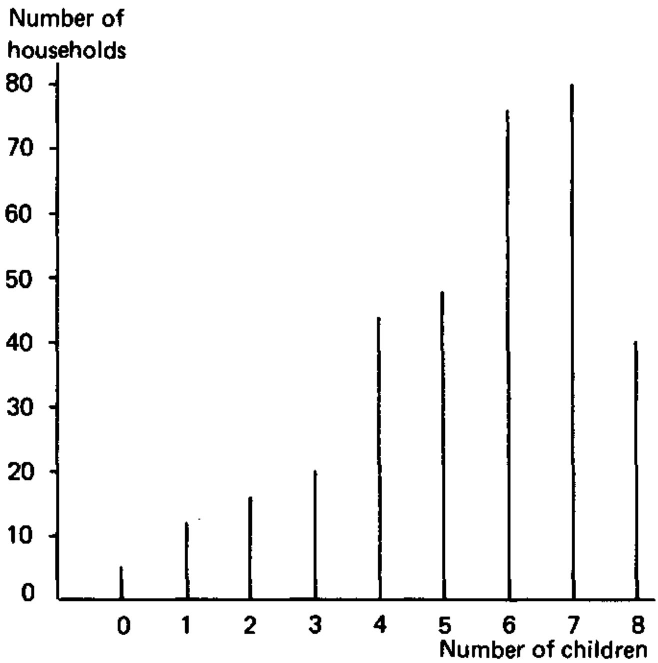


Figure 25.8. Needle graph of the number of children under the age of six in frequency of households with a specific village in 1976

25.4 Graphs of time series

A time series is a sequence of values taken by one variable over time (e.g. the price of cereals). Changes in the value of a variable over time can be presented on a graph, with time along the horizontal axis and the value along the vertical axis. A graphic presentation helps in visualizing how the variable changes over time and, without the need to use complex mathematical formulas, makes it possible to estimate what value that variable will have in the future.

Example of a graph of a time series

The price of millet at six markets was monitored at regular intervals throughout 1979. The average price per kilogram was calculated for each month for all the markets together. The results are presented in Table 25.6.

Table 25.6. Average price of millet (CFA/kg) in 1979 (combined data from six different markets)

Month	Price CFA/kg
January	56
February	61
March	65
April	68
May	62
June	63
July	65
August	66
September	64
October	58
November	49
December	45

The prices presented in a graph will show their fluctuations more clearly than a table will do. The graph is constructed as follows:

- Mark the months on the horizontal axis and the prices on the vertical axis;
- Mark the prices in Table 25.6 on the graph and connect the points with a line.

The resulting graph is shown in Figure 25.9.

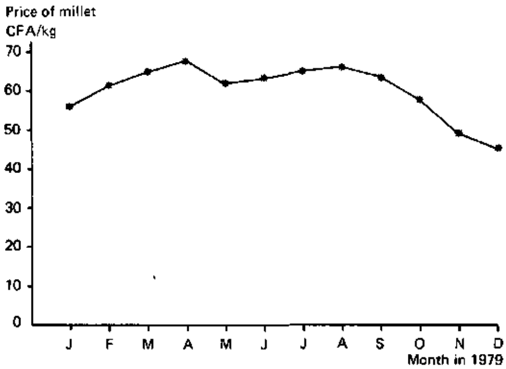


Figure 25.9. Time graph of the average price of millet in CFA/kg in 1979

When reading this graph, one should remember that the prices listed in Table 25.6 and marked as points on the graph correspond to one value only (average price each month). The lines between these points have no meaning, although the lines can sometimes be used as a tool for interpolation (i.e. estimating the value of a variable between two known points).

For example: If the figures for one month are missing, a line can be drawn between the price for the month before and the price for the month after. The missing price can be estimated as half-way between the two.

25.5 Diagrams

Diagrams are another kind of graphic presentation which are used if the number of classes is small. The advantage of diagrams is that they can

present variables that cannot be quantified (organizations, sex, etc.) and that they make a comparison over time possible.
Only two of the main types of diagrams will be discussed.

25.5.1 Bar diagrams

A bar diagram consists of a number of bars with the same sized bases, placed close together, but separated by a small space. The height of a bar and therefore its area represents the total frequency of the variable considered at one point in time or in one geographical location. The relative importance of each group within the variable is represented by its relative area within the bar. This makes it possible to compare the relative proportions of the groups and also to see how the frequencies change with time.

Example of a bar diagram

Table 25.7 presents the figures on cotton production in Francophone West African countries for the years 1973 to 1979. (Only the production marketed via the official government channels was considered.)

Table 25.7. Production of cotton, in tons, in six countries of West Africa 1973-1979 (Source: *Compagnie Française pour le Développement des Fibres Textiles*)

Country	Production in tons in year					
	1973/74	1974/75	1975/76	1976/77	1977/78	1978/79
Ivory Coast	58,465	59,939	65,041	75,413	102,929	114,886
Upper Volta	26,669	30,563	50,695	55,253	38,043	59,956
Mali	50,870	61,182	103,391	118,875	113,761	127,690
Niger	3,841	7,911	11,128	7,150	3,816	4,370
Senegal	32,584	42,099	30,685	45,208	37,166	33,806
Togo	8,784	10,648	9,772	7,063	4,537	14,365
Total	181,213	212,342	270,712	308,962	300,252	355,073

This table gives more detail than the average reader requires and it is difficult to read. The trends are more obvious if the data are presented in a bar diagram as in Figure 25.10.

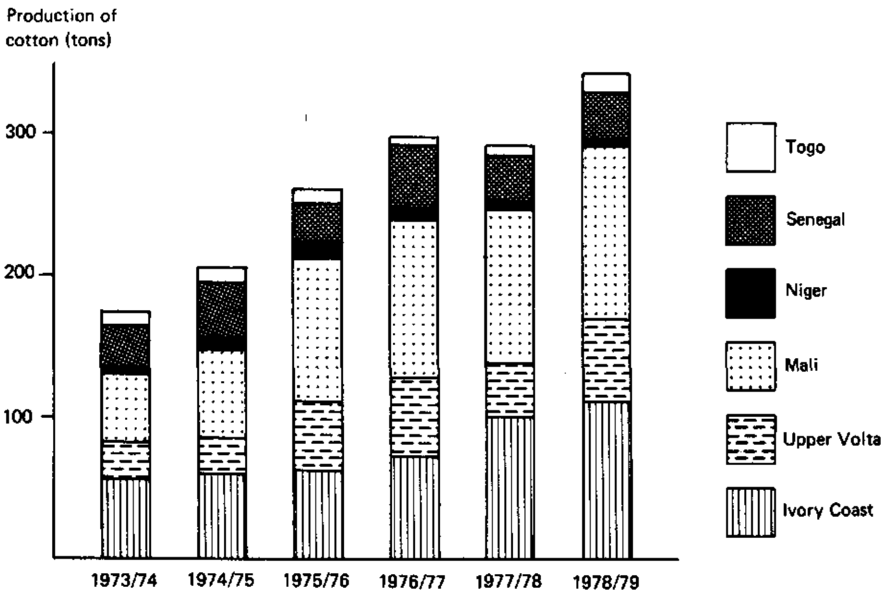


Figure 25.10. Bar diagram of the distribution of the cotton production in six countries of West Africa from 1973 to 1979

25.5.2 Circle diagrams

The information shown in a bar diagram can also be shown in a circle diagram, with a circle instead of a bar used for each year. A circle diagram is more difficult to draw but makes a report more attractive to the reader.

The cotton production in Table 25.7 and Figure 25.10 is presented in circle diagrams in Figure 25.11. The size of each circle is proportional to the total production for a given year. Each circle is divided into as many segments as there are countries, the segments being proportional to the production of each country.

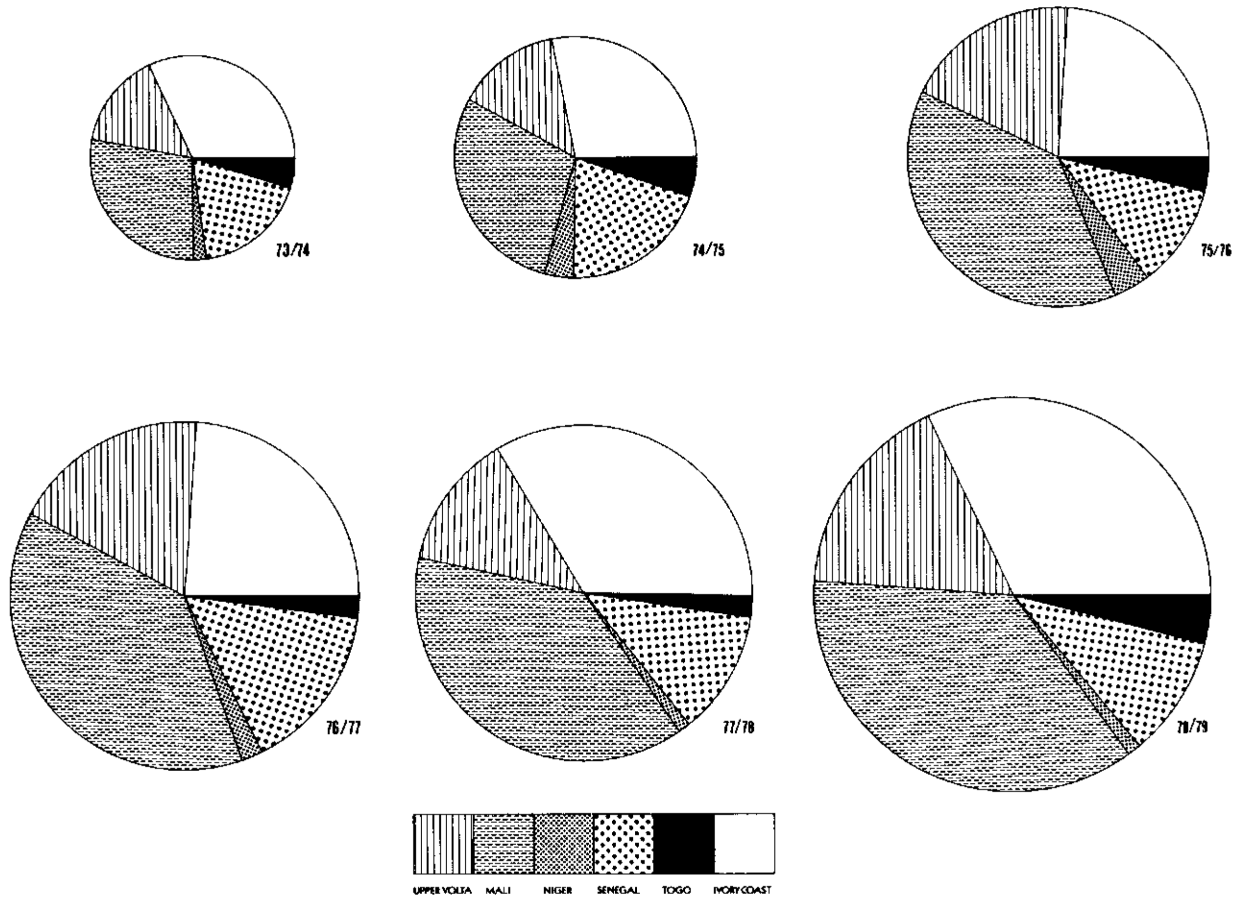


Figure 25.11. Circle diagrams of the distribution of cotton production in six countries of West Africa from 1973 to 1979

26 SUMMARIZING DATA

In Chapters 24 and 25, we showed how to present series of data in tables and graphs. In this chapter, we shall show how a series of data can be summarized by one or two values that can be used to represent it.

26.1 Central value

Normally, a series of data is clustered around a central value. This value can be used to represent the data. A central value provides much less information than the frequency distribution does, but it is a convenient way of summarizing the distribution. Just as there are different ways of presenting data in tables and graphs, there are also different ways of calculating a central value.

26.1.1 Arithmetic mean ($\bar{X}$)

The central value most often used is the arithmetic mean ($\bar{X}$). It is simply the total sum of all the values, divided by the total number of values. (This is what people often call the average.) The arithmetic mean can be used to summarize most data.
In mathematical symbols, a value can be represented by X and the total number of values by n; X_i represents each value in turn. The symbol Σ represents the process of adding up. The sum of all the values can thus be represented as:

$$\sum_{i=1}^{n} X_i$$

The arithmetic mean $\overline{X}$ is then calculated with the following formula, which shows the sum of all the values being divided by the total number of values:

$$\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Example of an arithmetic mean

If ten farmers harvested respectively 890, 1,200, 900, 750, 280, 690, 1,500, 1,490, 1,350, and 950 kg of cowpeas, the sum of all the productions is 10,000 kg and there are 10 values. The arithmetic mean is calculated as $\overline{X} = \frac{10,000}{10} = 1,000$ kg.

26.1.2 Weighted average ($\overline{X}_w$)

Although the arithmetic mean can be used for most data, it is sometimes better to follow a different approach. If all the data are not equally important, the weighted average summarizes the data better than the arithmetic mean. A weighted average is the arithmetic mean of values to which weights have been attached to make allowance for their relative importance. In a formula, it can be written as:

$$\overline{X}_w = \frac{\sum_{i=1}^{n} w_i X_i}{\sum_{i=1}^{n} w_i}$$

where

$\overline{X}_w$ = weighted average

w_i = weight of i

X_i = value of data number i

n = total number of values

Example of a weighted average

A farmer has three fields of different sizes, all cultivated with cowpeas. The fields have areas of 1.0 ha, 1.6 ha, and 0.4 ha. At the end of the cropping season, their yields were, respectively, 900, 600 and 1,200 kg/ha. The question is how to calculate the farmer's average yield of cowpeas. The arithmetic mean is $\frac{900 + 600 + 1{,}200}{3} = 900$ kg/ha, but it ignores the relative importance of the sizes of the fields, although the largest is four times larger than the smallest. To obtain the weighted average, weights are attached to the fields in accordance with their areas. The weighted average is thus the sum of the area of each field multiplied by its yield, divided by the sum of the areas. In other words, it is the farmer's total production of cowpeas divided by the total area cultivated:

$$\frac{1.0 \times 900 + 1.6 \times 600 + 0.4 \times 1{,}200}{1.0 + 1.6 + 0.4} = 780 \text{ kg/ha}$$

This example clearly illustrates the difference between the arithmetic mean and the weighted average. It also shows that the weighted average is a good way of summarizing the yields of one crop grown by a farmer on different fields, because it reduces the yield data to one per farmer.

The advantage of having the weighted average yield per farmer and how this value can be used will be illustrated in the next example.

Example of the use of the weighted average

Five farmers in a village were selected as being representative of that village. These farmers grew millet on a varying number of fields. The yields they obtained on each field are shown in Table 26.1.

Table 26.1. Millet yields obtained by five farmers

Field number	Farmer number 1		2		3		4		5	
	Area ha	Yield kg/ha	Area ha	Yield kg/ha	Area ha	Yield kg/ha	Area ha	Yield kg/ha	Area ha	Yield kg/ha
1	0.1	700	0.5	600	1.7	1,200	2.5	280	1.0	600
2	0.6	280	0.3	760			1.2	320	2.2	750
3	0.2	300	0.1	300			0.8	600	2.1	100
4	0.8	820	0.6	840					0.17	830
5			0.2	190						
6			0.3	200						

From this table it is possible to estimate the mean millet yield of the village. This is done in a two-stage procedure. First the weighted average yield for each of the five farmers is calculated. This gives 560, 580, 1,200, 350, and 510 kg/ha (rounded off). The second stage is to obtain the mean of these figures (640 kg/ha), which can be regarded as the mean millet yield of the village.

If an arithmetic mean of the yield of each field had been taken, the farmer with six fields would have been represented six times; the farmer with one field only once. Thus the farmer with six fields would have had a stronger influence on the mean than the farmer with only one field. If a weighted average of all the yields had been taken, Farmer 4, with 4.5 ha, would have been more strongly represented than Farmer 2, with 2 ha.

The weighted average can also be used when the frequency distribution of a variable is known, but not the original data. This happens when data from reports are used in which the data were presented in a frequency distribution in a table or graph. In this situation, the classes in a frequency distribution can be represented by their middles, while their weights are the respective frequencies. This weighted average can be used as an approximation of the arithmetic mean.

Example of the use of the weighted average as an approximation of the arithmetic mean

The frequency table of the cotton yields of forty farmers is presented in Table 26.2.

Table 26.2. Frequency distribution of the cotton yields of forty farmers

Cotton yield in kg/ha (classes)	Number of farmers (frequencies)
250 - < 500	5
500 - < 750	9
750 - < 1,000	14
1,000 - < 1,250	7
1,250 - < 1,500	3
1,500 - < 1,750	2
Total	40

From this table, the weighted average can be calculated:

$$X_w = \frac{5\times375 + 9\times625 + 14\times875 + 7\times1,125 + 3\times1,375 + 2\times1,625}{40}$$

$$= \frac{35,000}{40} = 875 \text{ kg/ha}$$

The weighted average can be calculated even more easily from the relative frequency distribution, with the middle of each class being multiplied by its percentage and the total sum being the weighted average.

26.1.3 Median (Me)

The arithmetic mean or the weighted average are almost always calculated. In most cases they give a satisfactory approximation and are useful tools for further analysis. If the central value is to be used as a final presentation of data, however, it is important to consider some other possible approaches. One of these is the median.

The median is the middle value of all the ranked values, the one that has as many values above it as below it. The median is not influenced by extreme values. This can be an advantage if a few extremes tend to disturb the arithmetic mean. Another advantage is that the median can be easily determined from a cumulative frequency histogram or polygon. It can also be used for a frequency distribution with an open first or last class (e.g. a yield of 1,200 kg/ha or more).
To identify the median, the values should first be ranked from smallest to largest. If two values fulfil the requirement of having as many values above them as below them, the median is the average of these two values.

Example of a median

The peanut yields of twelve farmers can be ranked from smallest to largest as follows:

640 700 770 950 1,070 1,090 | 1,460 1,460 1,480 1,560 1,580 1,670

Here, 1,090 could be the median, but 1,460 fits the requirement too, so the median is decided as $\frac{1,090 + 1,460}{2} = 1,275$ (average of the two numbers).

The median can be calculated from a frequency distribution by assuming that the data are distributed equally over the class in which the median is to be found.

Example of a median from a frequency distribution

A survey gathered income data from forty-nine small traders over a three-month period. The results are summarized in Table 26.3.

Table 26.3. Frequency distribution of the income from trade of a sample of forty-nine small traders over a three-month period

Income in CFA (classes)	Number of traders (frequencies)
0 - < 4,000	6
4,000 - < 8,000	16
8,000 - < 12,000	12
12,000 - < 16,000	7
16,000 - < 22,000	4
22,000 - < 30,000	3
30,000 and more	1
Total	49

As forty-nine incomes are being considered, the median is the twenty-fifth of the ranked incomes. This is in the class 8,000 - < 12,000 CFA because there are twenty-two incomes below that class and twelve in the class itself. If we assume that the twelve incomes in the 8,000 - < 12,000 CFA class are equally distributed over the class, the median is found to be the third of the ranked incomes in that class. It can be calculated as follows:

$$Me = 8{,}000 + \frac{3}{12} \times 4{,}000 = 9{,}000 \text{ CFA}$$

26.1.4 Mode (Mo)

The mode can be defined as the middle of the modal class, which is the class that contains the highest frequency (the greatest number of values). This measure of the central value depends on the exact choice of the classes and should therefore be used carefully, especially if the sample is small. The mode is appropriate for frequency distributions (e.g. income) in which there may be a few very extreme values.

Example of a mode

In the frequency distribution of the incomes from trade (Table 26.3), the modal class is Class 4,000 - < 8,000 CFA. The mode is 6,000 CFA (the middle of that class).

26.2 Measure of dispersion

A central value alone provides no information about the distribution of the values. Two distributions of the same variable can have the same mean although they are very different. Several ways of indicating the dispersion of the data (how the data are spread around the mean) will be discussed in this section.

26.2.1 Standard deviation (s)

The standard deviation (s) is one of the most widely used measures of dispersion. It indicates how closely the data are grouped around the mean. A small standard deviation in relation to its mean indicates that the data are grouped closely around the mean. (If all the observations were equal, the standard deviation would be zero.) A large standard deviation in relation to its mean indicates that the values are spread out in relation to the mean, so there are great differences between values. The standard deviation thus indicates how well the mean represents the data.
The standard deviation is a key element in many statistical formulas. Described in an equation, it is:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}}$$

where

X_i = value of data number i
$\overline{X}$ = mean of all values
n = total number of values

Because of the way in which the standard deviation is calculated, the values far from the mean influence the result more than the values close to it.

The standard deviation can still be calculated even if the original data are not available but only a frequency distribution. The standard deviation is then calculated using the weighted average calculated from the frequency distribution as the mean, the middles of the classes as the values, and the frequency of each class as the weight of each class middle.

$$s = \sqrt{\frac{\Sigma f_i(m_i - \overline{X}_w)^2}{\Sigma f_i}}$$

where

$\overline{X}_w$ = weighted average

f_i = frequency of Class i

m_i = middle of Class i

Examples of calculating the standard deviation

A survey was conducted to find out the size of the cattle herd of a sample of households. In one village, the heads of cattle per household for the six farmers interviewed were: 11, 27, 13, 2, 7, 72.

The mean herd size is 22. The standard deviation is:

$$s = \sqrt{\frac{(11-22)^2 + (27-22)^2 + (13-22)^2 + (2-22)^2 + (7-22)^2 + (72-22)^2}{6}}$$

$$= \sqrt{\frac{121 + 25 + 81 + 400 + 225 + 2,500}{6}} = 23.6 \text{ heads (rounded off)}$$

A standard deviation of 23.6 is large in relation to a mean of 22, which shows that the data are widely dispersed.

The standard deviation can also be calculated from a frequency table such as the frequency distribution of the cotton yields of forty farmers presented in Table 26.2. The weighted average was 875 kg/ha. The s can be calculated as follows:

$$s = \sqrt{\frac{5 \times (375-875)^2 + 9(625-875)^2 + 14(875-875)^2 + 7(,1125-875)^2 +}{40}}$$

$$\sqrt{\frac{+ 3(1,375-875)^2 + 2(1,625-875)^2}{40}} = \sqrt{\frac{1,250,000 + 562,500 + 0 +}{40}}$$

$$\sqrt{\frac{+ 437,500 + 750,000 + 1,125,000}{40}} = 321 \text{ kg/ha (rounded off)}$$

A standard deviation of 321 is small in relation to a mean of 875, which shows that the data are more closely grouped around the mean than in the previous example.

26.2.2 Range

Another indication of dispersion is the range, which is found by subtracting the lowest value from the highest value. A range can be useful when the data are not very reliable or when only a few data are available. It is easy to calculate but is not a good indicator if the two extreme values are exceptionally far from the rest. This should be checked carefully before a range is used.

Example of a range

The cotton yields in a village are 650, 820, 960, 1,100, 1,260, 1,300, 1,470, and 1,610 kg/ha. The range is 1,610 - 650 = 960 kg/ha.

26.3 Index numbers

The value of a variable often changes with time and so can be presented in a time series (Chapter 25, Section 5). A time series is useful to show changes in prices or in yields or production figures in a certain area. If one wants to compare such changes in different areas, however, one cannot simply compare their time series because they are likely to have different levels of values or are expressed in different (monetary) units. To make the series comparable, the data can be expressed in index

numbers usually on a yearly basis.
An index number is a ratio that expresses as a percentage the value of a variable for one year in relation to its value in another year, which is chosen as basis year. Any year can be chosen as basis year, but usually the first year in a series is chosen. It is given the value of 100, a number without a unit of measurement.

For example: The price of a bicycle in 1970 was 8,300 CFA and in 1978 it was 29,000. Taking 1970 as the basis year and giving it the value of 100, the index number for bicycle prices in 1978 is $\frac{29,000 \times 100}{8,300}$ = 350 (rounded off). The price has increased by roughly 350 per cent; it has multiplied by 3.50 (8,300 × 3.50 = 29,000, rounded off).

When two series of index numbers are to be compared, they must, of course, have the same year as basis.

Example of index numbers

The total cotton production in Niger and Upper Volta from 1961 to 1968 is listed in Table 26.4.

Table 26.4. Total cotton production in Niger and Upper Volta, in tons of seed cotton, from 1961 to 1968 (data from *Compagnie Française pour de Développement des Fibres Textiles* 1949/1969. p.40)

Year	Niger	Upper Volta
1961	2,320	2,352
1962	4,676	6,621
1963	5,457	8,048
1964	6,808	8,775
1965	6,075	7,463
1966	6,708	16,297
1967	6,169	17,274
1968	7,010	31,980

It is easier to compare the cotton production in these countries and the changes from year to year if index numbers are calculated. The year 1961 will be taken as basis year. (Another year can be taken as long as it is the same year for both countries.) The index for 1961 is thus 100 for both series. The production of 1962 can be represented by $\frac{4,676 \times 100}{2,320} = 202$ for Niger and $\frac{6,621 \times 100}{2,352} = 282$ for Upper Volta (both results rounded off). The index numbers in both series are calculated for each year.

Table 26.5. Cotton production in Niger and Upper Volta, in tons of seed cotton, from 1961 to 1968

	Niger	Upper Volta
Year	Production index (1961=100)	Production index (1961=100)
1961	100	100
1962	202	282
1963	235	342
1964	293	373
1965	262	317
1966	289	693
1967	266	734
1968	302	1,360

From this table it is easy to see that the cotton production increased much more rapidly in Upper Volta than in Niger, especially in the last three years of the series - something the original data did not show so clearly.

27 SIMPLE ANALYSES OF DATA

The last three chapters concentrated on the various ways of presenting and summarizing the data. Most monitoring units will also receive requests for more detailed analyses of the data. This chapter will describe how the results of two samples can be compared to see whether they come from the same population or from different ones; it will also explain the procedure for estimating relationships between variables. The analyses can all be done by hand or on a small calculator.

27.1 Comparing samples

Most of the samples that have been monitored will have been selected to represent different populations (e.g. farmers inside and outside the project) or strata of one population (e.g. farmers from different agroclimatological locations). Those requesting information on such populations or strata will want them to be compared to find out whether they differ. This cannot be shown by a difference in means only, as that difference can be due to random factors and to the sampling process. The distribution of the data indicated by the dispersion has to be taken into account as well.

In the chapter on sample selection (Chapter 17) it was stated that a sample should have at least ten to fifteen elements in each strata. This makes it possible to apply a relatively simple method to compare two samples and to draw conclusions about the two populations the samples were drawn from. The method tests whether the two samples were drawn from populations with the same mean and whether the populations can be considered to be the same. If the test indicates that this is not true, it can be concluded that the two samples are drawn from populations with different means.

The test procedure is as follows: Suppose there are two samples (e.g. the incomes of farmers in two villages). The first sample has n elements, called X_i (i = 1....n). The second sample has m elements, called Y_j (j = 1....m). For both samples, the mean ($\bar{X}$ and $\bar{Y}$) and the standard deviation (s_X and s_Y) can be calculated.

The mean and the standard deviation of both samples can be combined into a quotient z, which is defined as:

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{s_X^2/n + s_Y^2/n}}$$

This quotient will be smaller if the two means, $\bar{X}$ and $\bar{Y}$ are close to each other. As was shown in Chapter 26, the standard deviation is an indicator of the dispersion of a sample; the higher the standard deviation is, the more dispersed the sample is. It is logical that the more dispersed a sample is (and therefore also the population it comes from), the more the mean of each sample is influenced by the way it was selected; the difference between two sample means therefore has to be bigger to be statistically significant.

The conclusion drawn from a comparison of samples is never absolutely reliable. The reliability of the conclusion is indicated by the level of significance. There are tables that indicate the limits of z within which it can be concluded that the means of two samples are equal at different levels of significance. As a rule of thumb, if z is between -2 and +2, it can be concluded that the means of two samples are equal. Otherwise the sample means are significantly different.

Example of comparing samples

In Villages A and B, which have different sources of income, all income-generating activities were monitored throughout the year to find out whether the average income in one village differed significantly from that in the other. Originally twenty-five farmers were selected at random in each village but because some of the data sets were incomplete, not all could be used for the analysis. The final results were:

- In Village A: the usable sample was sixteen farmers, who had an average income of 110,000 CFA with a standard deviation of 32,000 CFA;
- In Village B: the usable sample was twenty farmers, who had an average income of 200,000 CFA with a standard deviation of 40,000 CFA.

To test whether the villages have different incomes or not, the quotient z is calculated. For simplicity, the information is put in units of 1,000 CFA.

$$z = \frac{\bar{X} - \bar{Y}}{\sqrt{s_X^2/n + s_Y^2/m}} = \frac{110-200}{\sqrt{(32)^2/16 + (40)^2/20}} = \frac{-90}{\sqrt{144}} = \frac{-90}{12} = -7.5$$

Here z = -7.5, which is smaller than -2. Therefore it can be concluded that the two samples have different means and come from different populations and that the two villages have significantly different incomes, seemingly due to their different activities.

The information obtained by comparing samples (and therefore populations) can be of great value to a project. Comparisons can point out the different groups in a population of farmers and can indicate, say, their potential for accepting new technologies. The indication of differences between groups of farmers, or the absence of it, can lead to project programs better adapted to the farmers' needs.

27.2 Time series

A time series is a series of values of the same variable over time, usually at regularly spaced intervals. In Chapter 25, it was shown how such data are best presented in a graph. Here the discussion will be focussed on how to extrapolate from such a time series, or in other words, how to estimate the trend that future values are likely to follow.

A simple example of a time series is the way a price changes with time, but a time series can also describe such matters as the development of agricultural production in a region, or the labour supply, and many other variables that change with time.

27.2.1 Graphic extrapolation

The analysis of a time series usually begins with a graphic presentation of the variable over time. This shows whether the series seems to follow a regular pattern or not, and whether a trend can be discerned. Sometimes a trend can be interrupted.

> For example: Oil prices on the world market shot upwards in 1973, so in a time series of world oil prices two separate trends would be noticed - one up to 1973 and the other one afterwards.

In a series of market prices of cereals at local markets in Africa, the price fluctuates within a year in another pattern than that from year to year, so to discover the long-term trend in such a series, one has to exclude the monthly (or daily) fluctuations.
A simple way to determine a trend is to prepare a graph with each value represented by a point and then to draw a line that stays as close as possible to every point. This is not a very precise method but it is fast and often gives a general idea of what is going on. From such a line it is possible to extrapolate the trend for the near future, reading graphically what value the line would represent, say, next year.

27.2.2 Moving average

A more precise method of estimating a trend is that of the moving average. It is especially useful for data that follow a cyclical pattern. A relatively simple method, it starts with determining how long one cycle lasts. For prices of cereals at local markets, for instance, assuming there is one agricultural season a year, the cycle lasts one year. The trend can therefore be estimated by the 12-monthly moving average.

To calculate the 12-monthly moving average, the arithmetic mean of twelve monthly values is calculated, first for Values 1 to 12, then for 2 to 13, 3 to 14 etc. This means that twelve months are always represented in the average, so monthly influences can be regarded as having been eliminated. The moving average indicates the trend of the variable, excluding the cyclical influences.

Example of a moving average

The market prices of sorghum at Market A are presented in Table 27.1. The average price for the first 12 months is 61 CFA. It is indicated midway in that year. The moving average for Months 2 to 13 is also 61. As can be seen in the table, the series of moving averages fluctuates much less than original series. This means that except for the seasonal fluctuations, the price level changes very little.

27.3 Linear correlation

Most agricultural projects promote farming techniques that are new to the farmers. Often introduced are crop varieties which have a high yield potential when combined with fertilizer use and proper cultivation methods. On trial plots at an agronomic research station, potential yields have been determined for various levels of fertilizer use. Under real farming conditions, however, fertilizer may not be applied at the right time or at the correct level, so the relation between fertilizer and yield identified on the trial plots does not hold at the farmers' level. Project managers are usually interested in this real relationship, or correlation, because fertilizer use plays an important role in many packages of technical innovations.
In statistics, a correlation problem can be described as the search for a relationship between variables. The most simple correlation is a linear correlation with two variables. One variable is supposed to be dependent on the other because of theoretical principles and it is estimated that their relationship can be approximated by a linear function. The linearity of the relationship seems a severe constraint,

Table 27.1. Monthly prices of sorghum over two years at Market A (CFA/kg)

Year	Month	Price (CFA/kg)	Moving average over 12 months (CFA/kg)
1978	January	59	
	February	60	
	March	66	
	April	62	
	May	60	
	June	69	61
	July	65	61
	August	62	61
	September	62	61
	October	64	61
	November	50	61
	December	53	61
1979	January	56	61
	February	61	62
	March	65	62
	April	68	61
	May	62	61
	June	63	60
	July	68	
	August	66	
	September	64	
	October	58	
	November	49	
	December	45	

but in practice many non-linear functions can be approached by one or more linear forms. As well, many non-linear relationships can be transformed into linear ones by a simple mathematical transformation of the variables.

The two variables between which a relationship is suspected can be represented by Y and X, Y being the dependent variable (e.g. yield) and

X the independent variable (e.g. fertilizer). Their estimated relationship does not mean that the use of fertilizer completely explains the variations in yields, only that it is an important factor. There are usually other factors (the timing of agricultural activities, rainfall distribution, and so on) which influence the yield, but the correlation looks only at the most important factor(s) and tries to quantify the relationship.

27.3.1 Graphic analysis

Before starting the estimation procedure for the exact relationship, one first makes a scattergram to see whether the expected relationship can be shown graphically.
A scattergram is a graphic presentation of all the values of the independent variable (X) measured along the horizontal axis and the corresponding values of the dependent variable (Y) measured along the vertical axis. Each pair of values is represented by a point. An example of a scattergram indicating the relation between fertilizer use and sorghum yield is shown in Figure 27.1.

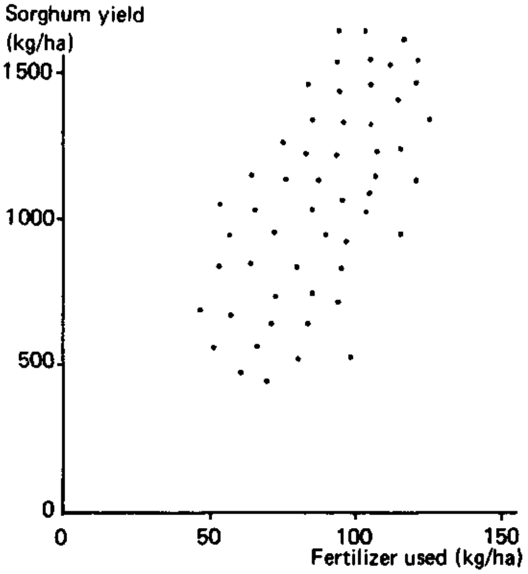


Figure 27.1. Scattergram of the relation between quantities of fertilizer used (kg/ha) and sorghum yields (kg/ha)

Just by looking at a scattergram, one can see whether a linear relation exists between the values of X and those of their corresponding Y. If X and Y have some relation, the cloud of points will be such that a line (not necessarily a straight line) can be drawn through them.

If the shape of the cloud of points is to be observed correctly, the scattergram has to be kept about square (range of X represented on about the same interval size as range of Y). Some possible clouds of points and their meanings will be explained below.

If there is perfect (linear) correlation, the points will be exactly aligned as shown in Figures 27.2 and 27.3. In both these figures, it is obvious that a linear relationship exists between the variables. The difference between the figures is that the first shows a positive correlation (if X increases, Y increases) whereas the second shows a negative correlation (if X increases, Y decreases).

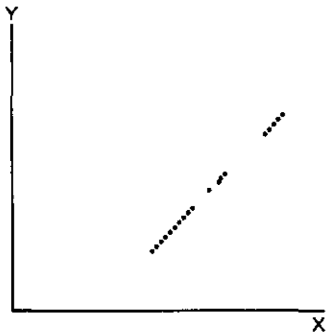


Figure 27.2. Scattergram of perfect positive correlation

Figure 27.3. Scattergram of perfect negative correlation

Perfect correlation is unlikely to occur with data from farm surveys, but the cloud of points can often confirm that there is a relationship. Figure 27.4, for instance, indicates a positive correlation and Figure 27.5 a negative correlation. Both these figures show that a linear relationship can be estimated to exist between the variables, but the scatter of points also indicates that there are other factors that influence and disturb their alignment.

Figure 27.4. Scattergram of positive correlation between two variables

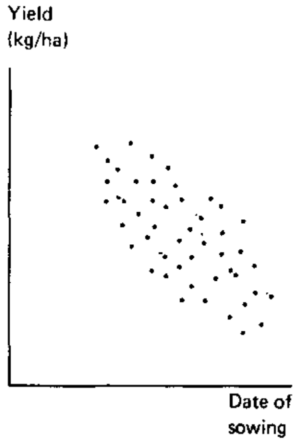


Figure 27.5. Scattergram of negative correlation between two variables

The absence of correlation is shown in Figures 27.6, 27.7, and 27.8. In the first, the cloud of points cannot (even with a great deal of imagination) be represented by a line, whereas in the other two the observations are aligned but with the same X or Y value. This shows perfect independence of the two variables.

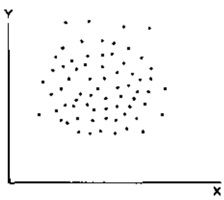
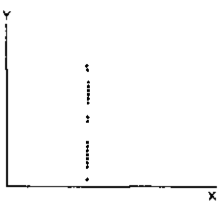
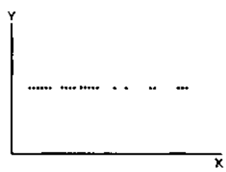


Figure 27.6. Scattergram of absence of correlation between two variables





Figures 27.7, 27.8. Scattergrams of two completely independent variables

A scattergram shows whether it is worthwhile or not to continue the estimation procedure. If no line (of whatever form) can be produced from the data (or part of the data), it is useless to continue since no relationship will be found, even though theoretically a relationship was expected.
A scattergram also shows which data fall completely out of line with the bulk of the data. Those data have to be checked again to make sure that they are not the result of errors. If no errors are found and the data indeed prove to be exceptional, they have to be excluded from the correlation procedure.
If the entire set of data shows no correlation, a better result may be obtained by looking only at part of the interval on the X axis.

27.3.2 Estimating the relationship

If the existence of a relationship is justified by the scattergram and also on theoretical grounds, the next step, is to estimate its mathematical equation. In general terms, the relationship can be estimated as:

$$Y = a + bX$$

where

X = independent variable

Y = dependent variable

a, b = parameters indicating the relationship between X and Y

The X in the above relationship is the main factor, but not the only one, which explains the value of Y: there are many other factors which are not considered. Therefore the relation $Y = a + bX$ is an approximation of reality, and the values will not exactly match the equation. The best possible approximation will be obtained by calculating the parameters a and b as follows:

$$b = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n} (X_i - \bar{X})^2}$$

$$a = \bar{Y} - b\bar{X}$$

where

X_i represents each X value in turn

$\bar{X}$ = mean of X_i

Y_i represents each Y value in turn

$\bar{Y}$ = mean of Y_i

n = number of paired values (X_i, Y_i)

It is important to check whether the approximation of the line $Y = a+bX$ to the points observed is good enough. This is done by calculating the correlation coefficient $r(X,Y)$.

27.3.3 Correlation coefficient

The correlation coefficient is calculated as follows:

$$r(X,Y) = \frac{\sum_{i=1}^{n} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2 \sum_{i=1}^{n} (Y_i - \bar{Y})^2}}$$

The correlation coefficient $r(X,Y)$ can take any value between -1 and +1. The higher the absolute value of r, the better the adaptation of the line to the observed values. The term $1-r^2(X,Y)$ indicates that part of the value of the dependent variable (Y) which is still not explained by the independent variable (X) in the relationship. An r of 0.5 does not mean that X explains 50 per cent of the value of Y; it only explains $1 - \sqrt{1-r^2(X,Y)} = 1 - \sqrt{0.75} = 1 - 0.87 = 13$ per cent. Perfect positive correlation, which means that an increase in the independent variable

(X) results in an increase of the dependent variable (Y), is indicated by r = 1. Perfect negative correlation, which means that an increase in X results in a decrease in Y, is indicated by r = -1.
As a rule of thumb, a correlation coefficient of 0.6 or less is considered too low to indicate the relationship.

Example of correlation

Throughout 1979, a monitoring unit has gathered data on agricultural income, area cultivated etc., from a sample of households. Preliminary analysis of the data has shown that a relationship exists between the area cultivated per worker and the total gross income per worker. In a country where extensive cultivation methods are applied, this is not illogical. The basic data available are tabulated in Table 27.2.

The first step in the correlation process is to make a scattergram of the data, as shown in Figure 27.9, to find out whether the data can be considered to have a relationship, and if so what kind of relationship.

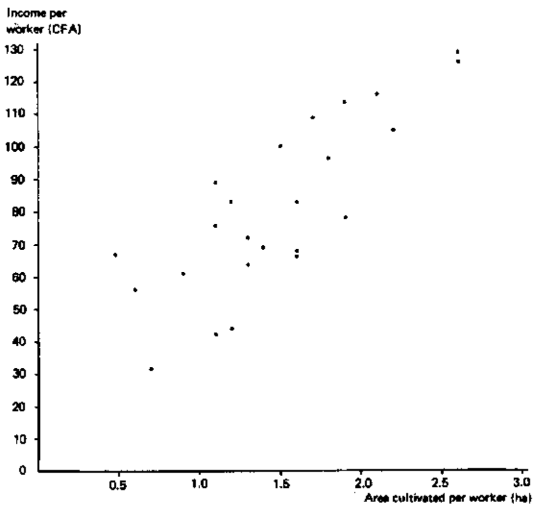


Figure 27.9. Scattergram of the relationship between the area cultivated per worker and the total gross income per worker

Table 27.2. Basic data on area cultivated per worker (ha) and total gross income from crop cultivation per worker (CFA) for 1979

Household number	Gross income per worker (x 1000 CFA)	Area cultivated per worker (ha)
1	116	2.1
2	126	2.6
3	56	0.6
4	32	0.7
5	105	2.2
6	72	1.3
7	78	1.9
8	83	1.2
9	100	1.5
10	67	1.0
11	66	1.6
12	109	1.7
13	113	1.9
14	76	1.1
15	64	1.3
16	68	1.6
17	69	1.4
18	129	2.6
19	96	1.8
20	44	1.2
21	42	1.1
22	89	1.1
23	83	1.6
24	61	0.9
n = 24	$\bar{Y}$ = 81	$\bar{X}$ = 1.5

This scattergram shows that it is worthwhile continuing with the procedure to estimate a linear relationship. Estimating the correlation equation Y = a + bX requires the calculation of a and b from the data of Table 27.3.

Table 27.3. Calculation table for parameters of the correlation problem of Table 27.2: Relation between income per worker (CFA) and area cultivated per worker (ha) for 1979

Household code	Y	X	$Y-\bar{Y}$	$X-\bar{X}$	$(Y-\bar{Y})(X-\bar{X})$	$(Y-Y)^2$	$(X-\bar{X})^2$
1	116	2.1	35	0.6	21.0	1,225	0.36
2	126	2.6	45	1.1	49.5	2,025	1.21
3	56	0.6	−25	−0.9	22.5	635	0.81
4	32	0.7	−49	−0.8	39.2	2,401	0.64
5	105	2.2	24	0.7	16.8	576	0.49
6	72	1.3	−9	−0.2	1.8	81	0.04
7	78	1.9	−3	0.4	−1.2	9	0.16
8	83	1.2	2	−0.3	−0.6	4	0.09
9	100	1.5	19	0	0	361	0.00
10	67	1.0	−14	−0.5	7.0	196	0.25
11	66	1.6	−15	0.1	−1.5	225	0.01
12	109	1.7	28	0.2	5.6	784	0.04
13	113	1.9	32	0.4	12.8	1,024	0.16
14	76	1.1	−5	−0.4	2.0	25	0.16
15	64	1.3	−17	−0.2	3.4	289	0.04
16	68	1.6	−13	0.1	−1.3	169	0.01
17	69	1.4	−12	−0.1	1.2	144	0.01
18	129	2.6	48	1.1	52.8	2,304	1.21
19	96	1.8	15	0.3	4.5	225	0.09
20	44	1.2	−37	−0.3	11.1	1,369	0.09
21	42	1.1	−39	−0.4	15.6	1,521	0.16
22	89	1.1	8	−0.4	−3.2	64	0.16
23	82	1.6	2	0.1	0.2	4	0.01
24	61	0.9	−20	−0.6	12.0	400	0.36
Subtotal			258−258	5.1−5.1	279.0−7.8	16,050	6.56
Total		1,944	36.0	0	0	271.2	
Mean		81	1.5				
Notation	$\bar{Y}$	$\bar{X}$	$\Sigma(Y-\bar{Y})$	$\Sigma(X-\bar{X})$	$\Sigma(Y-\bar{Y})(X-\bar{X})$	$\Sigma(Y-\bar{Y})^2$	$\Sigma(X-\bar{X})^2$

Estimating the b of the correlation equation can be done by:

$$b = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\Sigma(X - \bar{X})^2} = \frac{271.2}{6.56} = 41.34$$

and the a by:

$$a = \bar{Y} - b\bar{X} = 81 - 41.34 \times 1.5 = 18.99$$

The equation is therefore $Y = 18.99 + 41.34X$ or rounded off $Y = 19 + 41X$.

How well this equation fits the data is expressed by the correlation coefficient $r(X,Y) = \frac{\Sigma(X - \bar{X})(Y - \bar{Y})}{\sqrt{\Sigma(X - \bar{X})^2\Sigma(Y - \bar{Y})^2}} = \frac{271.2}{\sqrt{6.56 \times 16{,}050}} = 0.84.$

This r is higher than 0.6. Therefore it can be concluded that the relationship exists. This correlation coefficient explains $1-\sqrt{1-r^2}$ = 46 per cent of the values, which is rather high for farm data.

It can be concluded that there is a significant relationship between the area cultivated per worker and the total gross income per worker. This finding can therefore be used to predict the income per worker.

Suppose a worker cultivates 2 ha. What will his income be in 1979? This can be found by filling in the relation:

$Y = a + bX$

$Y = 19 + 41 \times 2 = 101$

His income can be predicted at 101,000 CFA.

27.4 Conclusions

Estimates of relationships are frequently calculated when farm data are being analyzed. Used properly, a linear correlation can reveal the importance of key factors in the achievements of the farmers. One should, however, use this relationship with great care, taking the following into consideration:

- Farm data are usually fairly imprecise, so relationships are often hard to prove, although this does not always mean that they are absent. On the other hand, it can happen that a relationship which was not expected is revealed. Such a relationship should be examined carefully to find the logic behind it because the existence of a relationship cannot be proved merely on the basis of a correlation function;

- Any correlation calculation will result in a relationship, but only the correlation coefficient will indicate whether the relationship is significant or not.

Only the most simple form of correlation has been presented here because it is the one most often used in farm data analysis and because its calculation can be done, without special programs, on most calculators. More complex methods include more than two variables, or even non-linear forms of correlation. The inclusion of more variables in the correlation may seem attractive, but they will greatly complicate the calculations and may lead to a misinterpretation of the results.

BIBLIOGRAPHY

Anthony, K.R.M., B.F. Johnston, W.O. Jones and V.C. Uchendu, 1979. Agricultural change in tropical Africa. Cornell University Press, Ithaca.

Barnum, H.N. and L. Squire, 1979. A model of an agricultural household - theory and evidence. World Bank Occasional Paper 27. John Hopkins Press, Baltimore.

Bradfield, D.J., 1970. Guide pour la formation des vulgarisateurs. FAO, Rome.

Centre d'Etudes et d'Experimentation du Machinisme Agricole Tropical, 1972. The employment of draught animals in agriculture. FAO, Rome.

Cernea, M. and B.J. Teffing, 1977. A system for monitoring and evaluation of agricultural extension projects. World Bank Staff Working Paper 272. World Bank, Washington.

Cernea, M.M., 1979. Measuring project impact: monitoring and evaluation in the Pider rural development project, Mexico. World Bank Staff Working Paper 332. World Bank, Washington.

Chambers, R., 1980. Shortcut methods in information gathering for rural development projects. Paper for World Bank Agriculture Section Symposium, January 1980.

Chatfield, C., 1949. Food consumption tables for international use. FAO Nutritional Studies 3. Washington.

Chou, M., D.P. Harmon Jr., H. Kahn and S.H. Wittwer, 1977. World food prospects and agricultural potential. Praeger, New York.

Clark, C. and M. Haswell, 1964. The economics of subsistence agriculture. Macmillan, London.

Clayton, E. and F. Petry (Eds.), 1981. Monitoring systems for agricultural and rural development projects. FAO Economic and Social Development Paper 12. FAO, Rome.

Collinson, M., 1972. Farm management in peasant agriculture: a handbook for rural development planning in Africa. Praeger, New York.

Couty, Ph., I.Y. Marchal, P. Pelissier, 1979. Maitrise de l'espace agraire et développement en Afrique tropicale - logique paysanne et rationalité technique. Actes de colluque de Ouagadougou, 4-8 December 1978. ORSTOM, Paris.

Deboeck, G. and W. Kinsey, 1980. Managing information for rural development - lessons from Eastern Africa. World Bank Staff Working Paper 379. World Bank, Washington.

Deboeck, G. and R. Ng, 1980. Monitoring rural development in East Africa. World Bank Staff Working Paper 439. World Bank, Washington.

Discar, W.J. and F.J. Massey Jr., 1969. Introduction to statistical analysis. 3rd ed. McGraw-Hill, New York.

Food and Agricultural Organization, 1970. Fertilizers and their use: a pocket guide for extension officers. 2nd ed. FAO, Rome.

Food and Agricultural Organization, 1974. Shifting cultivation and soil conservation in Africa. Papers presented at the FAO/SIDA/AECN Regional Seminar on Shifting Cultivation and Soil Conservation in Africa, Ibadan, Nigeria, 2-21 July 1973. Soils Bulleting 24. FAO, Rome.

Food and Agricultural Organization, 1977. Rapport sur le recensement mondial de l'agriculture de 1970. FAO, Rome.

Freeman, H.E., P.H. Rossi and S.R. Wright, 1979. Evaluating social projects in developing countries. OECD, Paris.

Hall, A.E., G.H. Cannell and M.W. Lawton, 1979. Agriculture in semi-arid environments. Ecological Studies Vol. 34. Springer, Berlin.

Hildebrand, P.E., 1979. Summary of the Sondeo Methodology used by ICTA (Instituta de Ciencia y Technologica Agricolos). Ministra de Agricultura, Guatemala. Guatemala City.

Hunt, K.E., 1970. Agricultural statistics for developing countries. Oxford University Press, Oxford.

Hursh, G., 1976. Third world surveys. Macmillan, London.

Imboden, N., 1978. A management approach to project appraisal and evaluation. OECD, Paris.

Imboden, N., 1980. Managing information for rural development projects. OECD, Paris.

Institut National de la Statistique et des Études Économiques (INSEE), 1961. Manual pour le formation d'enquêteurs dans de cadre une étude par sondage de budgets familiaux et du insommation dans un pays en

voi de développement. Ministère de la Cooperation, Paris.
Institut National de la Statistique et des Études Economiques (INSEE), 1962. Manual d'enquête agricola. 2nd ed. Ministère de la Cooperation, Paris.
Institute of Development Studies, 1979. Rapid rural appraisal. Papers presented at a conference held 4-7 December at the Institute of Development Studies, University of Sussex.
International Crops Research Institute for the Semi-arid Tropics, 1980. Proceedings of the International Workshop on Socio-economic Constraints to Development of Semi-arid Tropical Agriculture, 19-23 February 1979, India. ICRISAT, Andhra Pradesh.
Jodha, N.S., 1976. Considerations in measurements of rural household incomes. ICRISAT, Andhra Pradesh.
Johnston, B.F., 1958. The staple food economics of Western Africa. Stanford University Press, Stanford.
Hageboeck, M., 1979. Manager's guide to data collection. Office of Evaluation. Agency Int. Dev., Washington.
Kearl, B., 1976. Field data collection in the social sciences; experiences in Africa and the Middle East. Agricultural Development Council, New York.
Kowal, J.M. and A.H. Kassam, 1978. Agricultural ecology of savannah: a study of West Africa. Clarendon Press, Oxford.
Latham, M.C., 1972. Planning and evaluation of applied nutrition programmes. Nutritional Studies 26. FAO, Rome.
Little, I.M.D. and J.A. Mirrlees, 1974. Project appraisal and planning for developing countries. Heinemann, London.
Matlon, P., 1977. Farm level data collection in Northern Nigeria: guidelines for questionnaire design. Michigan State University, Michigan.
McLoughlin, P.F.M. (Ed.), 1970. African food production systems: cases and theory. John Hopkins Press, Baltimore.
Mellor, J.W., 1966. The economics of agricultural development. Cornell University Press, Ithaca.
Mood, A. and F. Graybill, 1963. Introduction to the theory of statistics. 2nd ed. McGraw-Hill, New York.
Nelson, M., 1974. The development of tropical lands. John Hopkins Press, Baltimore.
Norman, D.W., 1970. Labor inputs of farmers: a case study of the Zaria

Province of the North Central State of Nigeria. Samaru Research Bulletin 116, Samaru.

Norman, D.W., 1967, 1972. An economic study of three villages in Zaria province. Samaru Miscellaneous Papers 19, 37, 38, Samaru.

Norman, D.W., 1973. Economic analysis of agricultural production and labor utilization among the Haussa in the north of Nigeria. African Rural Employment Paper 4, East Lansing.

Norman, D.W., 1978. Methodology and problems of farm management investigation: experiences from Northern Nigeria. African Rural Employment Paper 8. Michigan State University, Michigan.

Norman, D.W., J.C. Fine, A.D. Goddard, J. Kroeker and D.H. Pryor, 1976. A socio-economic survey of three villages in the Sok close-settled zone. Samaru Miscellaneous Paper 65. Ahmadu Bello University, Zaria.

Payne, S.L., 1951. The art of asking questions. Princeton University Press, Princeton.

Perrin, R.K., D.L. Winkelman and E.R. Moscardi, 1976. From agronomic data to farmer recommendations: an economic training manual. Information Bulletin 27. International Centre for Corn and Wheat Improvement. CIMMYT, Mexico.

Perrin, R.K. and D.L. Winkelmann, 1976. Impediments to technical progress on small versus large farms. International Centre for Corn and Wheat Improvement (CIMMYT), Mexico.

Phillips, J., 1959. Agriculture and ecology in Africa. Faber and Faber, London.

Piclet, G., 1973. Notions d'économic générale et d'économic rurale. FAO, Rome.

Reh, E., 1976. Manual on household food consumption surveys. FAO, Rome.

Rossi, P.H. and W. Williams (Eds.), 1974. Evaluating social programs: theory, practices, politics. Seminar Press, New York.

Ruthenberg, H., 1971. Farming systems in the tropics. Clarendon Press, Oxford.

Scott, W., 1978. Measurement and analysis of progress at the local level. UNRISD, Geneva.

Stevens, R.D. (Ed.), 1977. Tradition and dynamism in small-farm agriculture: economic studies in Asia, Africa and Latin America. Iowa State University Press, Ames.

Struening, E.L. and M. Guttenlag (Eds.), 1975. Handbook of evaluation research. Sage, London.

United Nations. Dept. of Economic and Social Affairs, 1978. Systematic monitoring and evaluation of integrated development programs. U.N. Washington.

Upton, M., 1973. Farm management in Africa: the principles of production and planning. Oxford University Press, Oxford.

World Bank, 1981. World Development Report 1980. World Bank, Washington.

World Bank, 1981. Accelerated development in sub-Saharan Africa: an agenda for action. World Bank, Washington.

Wrigley, G., 1969. Tropical agriculture: the development of production. Faber and Faber, London.

Yang, H.P., 1968. Fact-finding with rural people. Agricultural Development Paper 52. FAO, Rome.

Yang, W.Y., 1965. Methods of farm management investigations. FAO, Rome.

Yule, G.U. and M.G. Kendall, 1965. An introduction to the theory of statistics. 14th ed. Griffin and Co., London.

Zarkovitch, S.S., 1965. Estimation of areas in agricultural statistics. FAO, Rome.

Zarkovitch, S.S., 1966. Quality of statistical data. FAO, Rome.