PN-AAP-105
ISN 33071

B-23

DRAFT FOR COMMENT
(1979)

Design and Management of Survey Research:

A Guide for Agricultural Researchers

Richard Bernsten
CRIA/IRRI Cooperative Program
P.O. Box 107
Bogor, Indonesia

TABLE OF CONTENTS

		Page
	Preface	1
I.	Introduction	2
II.	Why Conduct a Survey	3
III.	Focusing Your Research	5
IV.	Sampling Design	7
	A. The Sampling Unit	7
	B. Sampling Method	8
	C. Choosing Between Sampling Methods	12
	D. Sample Size	13
	E. Sampling Frames - Hidden Biases	13
V.	Primary Data Collection Instruments	15
	A. Village leader interviewer	17
	B. Group Interview	18
	C. Structural Individual Interview	19
	D. Farm Record Keeping	20
	E. Verification Measurements	22
	F. Choosing the Appropriate Primary Data Collection Instrument	24
VI.	Survey Questionnaire Development	24
	A. Questionnaire Design	25
	B. Wording of Individual Sections	27
	C. Format	34
	D. Operations Manual	39
	E. Translating	40
VII.	Pre-testing the Draft Questionnaire	41
	A. Personnel	41
	B. Location	41
	C. Indicators of Problem Questions	42
	D. Indicators of Problem Formal	42
	E. Inadequate Procoding	43

Page

VIII. Selecting Enumerators 43
A. Characteristics of Good Enumerators 43
B. People who Make Good Interviewers 44
C. Selecting Among Candidates 45
D. Reducing Staff Turnover 45

IX. Training Interviewers 46
A. Length of Training Program 46
B. Focus of the Training Program 47
C. Level of Supervision 48

X. Implementing the Field Survey 49
A. Obtaining Cooperation from the Village Leaders 49
B. Obtaining Cooperation From The Village Farmers 49
C. Supervision of the Enumerators 51
D. Special Problems in Field Implementation 53

XI. Choosing the Data Analysis System 55
A. Tabular sheets 56
B. Sorting strips 57
C. Computer Analysis 60

XII. Brief Introduction to Computer Analysis 62
A. Hardware 62
B. Softwear 63
C. The Statistical Package for the Social Sciences 65

XIII. Conclusion 78
References 79

P r e f a c e

When first attempting to conduct survey research several years ago, I found very few useful articles or books to guide my floundering. The available material tended to be abstract and did not focus on the types of problems actually confronted in trying to reconcile theory with practice. While some of this void has been filled during the past ten years, there is still a need for additional contributions. My motivation for writing this guide is to provide a practical reference for agricultural field workers wishing to conduct survey research, often with no background in social science. In retrospect, this is the type of booklet I would have liked to have had available several years ago before conducting my first survey.

The ideas in this guide are mainly the product of experience acquired over the past ten years while participating in various survey research projects. In addition, I have benfited from exposure to the ideas and experiences of other social scientists, both through personal contact and written material.

I. Introduction

In developing countries where secondary data is characteristically scarce, the survey can serve as a valuable tool for collecting information about the physical, social, institutional and economic environment. Through the careful analysis of this data, the survey researcher can make a significant contribution towards improving the quality of the information base upon which development policies and biological research priorities are formulated.

Recent developments in computer technology have permitted the survey researcher to rapidly analyze large volumes of data utilizing sophisticated multivariate techniques. Yet, the researcher must keep in mind that the quality of data collected is as - perhaps more important - to its ultimate usefulness as the number of questions asked or sample size.

Survey research is often viewed with suspicion by non-social scientists who are used to working with extremely accurate experimental data. While surveys can seldom generate such precise data, this is no reason for rejecting their value as research tools. We must not forget that the relevance and accuracy of experimental data depends on the care taken to design and conduct the trial. An unskilled field assistant can easily ruin a fertilizer field trial. Yet, it would be absurd to conclude that the random block is an inappropriate method for evaluating yield response to fertilizer. This is equally true for survey research. In order to generate meaningful data that can be used to develop agricultural policy guidelines, the researcher must carefully identify relevant hypotheses; design questions that will shed light on these issues; outline a strategy for field implementation; design a good data collection instrument; train the interviewers to carefully conduct the interview; and thoroughly analyze the resulting data.

In addition to improving the accuracy of data collected, the application of survey research methodology also increases the efficiency of primary data collection.[1] This is a very practical goal since survey research is quite expensive to conduct and the cost rises as the sample size increases. Consequently, substantial cost reduction can be achieved by the application of modern survey techniques.

---

[1] Efficiency - Cost per interview or information per interview (CYMMIT, 1978).

Finally, it is possible to successfully analyze survey data by a variety of systems — ranging from using the hand held electronic calculator to the IBM370 computer. Since available resources and analytical goals vary between projects, countries, and institutions within each country there is no single best approach. Each researcher must evaluate the alternative systems and develop a strategy most appropriate for the particular situation.

This booklet attempts to outline practical suggestions for:

- determining when a farm survey is required
- identifying the goals of your research so that the survey will generate data that can be used to answer the intended question
- choosing the appropriate sampling strategy
- deciding on the type of survey necessary
- designing the questionnaire
- selecting and training enumerators
- implementing the survey
- choosing between alternative data analysis systems

II. Why Conduct a Survey?

The first step in beginning a research project is to outline the general purpose of the proposed investigation. The more precisely this is done, the more purposefully you can proceed to review existing studies and secondary data.

A review of existing literature is necessary to determine what research has already been completed on the proposed topic in order to avoid duplicating studies which answer the same questions which you intend to explore. In the developing world, few studies find their way into the standard academic journals. Consequently, the best way to find out what research has been done is to visit the research institutions in your country. The university department in your field of interest (economics, sociology, agricultural economics, agronomy) may be able to assist. Additional possibilities include the Ministries of Agriculture, Public Works (Irrigation), Planning, and Bureau of Statistics. In some countries, a National Research Board will coordinate all research activities in the country, and should be contacted. Also, a great deal of research is

funded by international organizations and foundations. Hence, staff members at the offices of the Rockefeller and Ford Foundation, Agricultural Development Council, international agricultural research institutes (IRRI, CIAT, ICRISAT, IITA, AVRDC), IDRC (International Development and Research Center, Canada), USAID, CIDA (Canadian AID), FAO, and World Bank may have knowledge of studies related to the one you are planning.

It may be possible to complete the proposed research by analyzing available secondary data that can be found in existing population and agricultural censuses. In addition, field data collected by other researchers can often be obtained, and re-examined in light of your particular research problem. When such resources are available, the researchers should seriously consider using these materials instead of conducting their own field investigation.

In addition to published (and collected) secondary data, in many countries village level secondary data can be collected quite easily. For example, most Indonesian village "statistics offices" maintain data on population by age and sex; hectares in various land use classifications (irrigated, rainfed, upland, home gardens, estate crops etc.); number of farms by size class; hectares planted and harvested of each crop by month; output and yield by crop; hectares enrolled in the government's intensification programs (Bimas and Inmas); yield and area in modern and local rice varieties; and number of tractors and carabou. In addition, local officials can provide information about such agro-support services as cooperatives, seed farms, extension workers and rice mills.

While some of these statistics are not always completely accurate, the researcher can use them to identify basic village characteristics. For example, an excessive proportion of females in the population may indicate extensive outmigration of male labor; hectares per farmer indicate the existing pressure on land; percent of farms in each size class suggest the distribution of wealth; planting and harvesting statistics show the area double cropped to rice and probable cropping patterns; low yield indicates the existence of production constraints; hectares per tractor and hectares per carabou suggest the level of land preparation mechanization; hectares per extension agent and percent of the area in the intensification program suggest the adequacy of agro-support coverage. Consequently, in countries

where such a rich set of village level secondary data is routinely collected through existing institutions and this data is adequate for the proposed research goals, it may be unnecessary to conduct a farm survey.

Even if a thorough search of the literature uncovers no studies specifically in your proposed research area; and no available secondary data is found which can be used for analyzing the proposed research issues; this effort has not been wasted. First, familiarizing yourself with studies generally related to the proposed research will help you to more carefully focus your research on important issues which have been neglected previously. Second, a literature review will help you to avoid some of the methodological weaknesses which may characterize earlier studies. Third, the collection and evaluation of secondary data will assist in providing an overview of the general agricultural situation. Finally, a farm survey is only a means to an end and not an end in itself. Only after reviewing both previous studies and secondary data can you be sure that a survey is actually necessary to achieve your research goals.

III. Focusing Your Research

The first step in developing a survey is to identify the analytical goals of the research. Is the intent to explore, describe, or explain relationships? Once this is done, it will be possible to make certain decisions regarding implementation (Kearl, 1976 in Norman 1977). For example, if the purpose of the study is to simply explore (i.e. provide basic insights and generate hypotheses) non-random sampling may be adequate. On the other hand, if the researcher wishes to describe the sample in order to generalize about the population, scientific sampling will be required. If the study proposes to explain relationships, multivariate techniques must be used. In the latter two cases, in order to minimize sampling error,[2] sample size must be large.

In farm level research, it is also necessary to identify the disciplinary focus of the research. If we are only interested in purely economic issues, the economist is adequately trained to develop the

---

[2] Sampling error - The difference between the true value of the population parameter and the value estimated from the data collected.

survey with only minimum consultation with biological scientists. Yet, this is seldom the case. Typically farm level surveys deal with interdisciplinary issues that can only be successfully evaluated when both economists and biological scientists cooperate closely. Interaction is particularly important in the early stage of research and questionnaire design. For example, if we wish to include questions to permit site description, the agronomists should be consulted to identify the critical agro-climatic factors that will influence cropping patterns and yield. In the same vein, if we intend to evaluate new technology, it is necessary to consult with the biological scientists to determine exactly which inputs and management practices should be monitored. The interdisciplinary approach not only increases the usefulness of survey research to biological scientists by insuring that the critical agricultural factors will be covered in the survey, but also the credibility of the results is increased when the biological scientist actively participates in designing the research.

Once the analytical purpose and disciplinary scope is decided, the researcher must sharply focus the purpose of the study. Unless this is done it is likely that after the data is collected, the researcher will realize that the questions necessary for obtaining data to test important hypotheses were not asked. Such a dilemma can be best avoided by first writing out each of the issues to be studied. For example, "Do farmers benefit from planting modern varieties?" Second, each issue should be stated as a formal hypothesis. In the above case the null hypothesis might be "the net returns of farmers planting modern varieties is equal to the net returns of farmers not planting modern varieties" (i.e. Ho : $\bar{X}_{MV_{net}} = \bar{X}_{TV_{net}}$). Third, the data requirements to test this hypothesis such as farmer's output, planted area, inputs, and the product price should be listed in order to insure that the questionnaire will include items to measure all of these factors, unless available secondary data can be relied upon to supply some of this information. Fourth, once each research issue has been precisely identified, the hypothesis explicitly stated, and data requirements listed, it is advisable to draft the dummy tables into which you plan to present the data in the final report. This step serves as a check to determine if you have actually planned to

collect the data required to write the proposed research report. Finally, where computations are to be made (net returns, etc.), it is useful to actually generate a set of hypothetical data. These values can be used to make the desired calculations in order to determine if the interview schedule includes all questions required to make the intended computations (Lynch, 1978).

A useful way to organize these steps in questionnaire development is to first list all of the issues to be covered in the questionnaire. Next, reorder these issues into logical subject groupings. For example, all issues related to production, consumption, marketing, etc. may be listed on separate pages in a 8" x 11" loose leaf binder. Third, write out all of the hypotheses dealing with each issue and place these on a sheet of paper in the appropriate section. Fourth, draft the questions to be asked to collect the required data for testing each hypothesis and place these pages behind the respective hypothesis. Finally, draft all of the dummy tables into which the data will be put and place these at the end of the relevant section. This organization of the questionnaire development process permits rapid reference between questionnaire sections and is a handy way to keep all of the pieces of the survey together.

IV. Sampling Design

If the characteristics of all members of a population were exactly the same, there would be no need to worry about sampling procedure. We would only need to select one individual in order to identify the population characteristics. For example, the chemist can study the properties of one atom of gold and know that his observations hold true for all samples of this metal. In contrast, the diversity of our social environment requires that we sample several members of the population for evaluation, before any conclusions can be drawn.

A. The Sampling Unit

Before deciding on a sampling design, it is necessary to define the sampling unit. Choice of the sampling unit depends on the purpose of the study. Do the research issues suggest that we sample fields, farmers, farm families, cooperative members, or extension workers? Farm level research usually focuses on the farm family. Yet, the researcher must

still decide on an acceptable definition of this unit. For example, several members of one family may live in the same house — each farming a different piece of land, but cooperating in decision making and exchanging labor (Norman, 1976). The researcher must decide if — for his purposes — this organization constitutes a single or several sampling units. Once the sampling unit has been identified, it will be easier to finalize your sampling design.

B. Sampling Methods[3]

The purpose of sampling is to select a sub-set of the population which has the same characteristics of the whole group or is representative of the population.[4] If this goal is met, we can make valid inferences about the population without interviewing every member. Consequently, since only a small portion of the population must be interviewed, survey costs and management problems are reduced. Several sampling procedures are available, including non-random, random, systematic, stratified, quota, cluster and multi-stage techniques.

1) Non-random sampling is accomplished by purposefully selecting sampling units without the use of a population list. An example of this method would be to interview all farmers you meet in the morning as you walk through a village. While this method may be appropriate for generating hypotheses, it is impossible to generalize your results to the population of farmers. Since sampling was not random,[5] you have no way of knowing if the individuals interviewed were representative of the population. In fact, farmers who are not working in their farms in the morning may cultivate very small holdings, hold non-farm jobs, or be lazy. Consequently, it would be very misleading to generalize about the population from this non-random sample.

---

[3]Discussion of sampling design based on Pareh, Cristina et.al. 1973.

[4]Population - All of the elements (such as farmers, households etc.) from which the sample is actually selected

[5]Random - A sampling procedure which insures that each respondent has an equal chance of being selected.

2) Simple random sampling requires a list of the population (sampling frame). From this list, a predetermined number of units are selected in a manner which will insure that each unit has an equal chance to be chosen. The random number table is a frequently used aid in sample selection.

Each unit is numbered 1 through N (number of units in the population). Then the researcher randomly selects a starting number on the table and finds the population unit with the corresponding number. The researcher then moves across the table, selecting the next number and the respective population unit. This procedure is followed until the desired sample size is obtained. The major advantage of random sampling is that it is easy to implement. It is also appropriate if the population is concentrated in a single confined area. On the other hand, it may be difficult to obtain a population list, the selected units may be geographically dispersed and therefore expensive to reach, and when the population characteristics are very variable, it provides imprecise[6] estimates.

3) Systematic sampling is similar to random sampling. A random starting point is selected and every K^{th} (for example, every tenth unit) unit is chosen. If the population consists of 300 farmers and you want to select 20 respondents, this means one in every 15 units must be selected to obtain the required total. A number between 1 and 15 is chosen and that unit is selected. Then, every 15th unit after the first one selected is also chosen as part of the sample. The major advantage of systematic sampling is that it is easier to carry out than random sampling. On the other hand, if the population list is not in random order (e.g. farmers listed in order of farm size) it should not be used because the resulting sample may be biased.[7]

4) Stratified random sampling is used when the researcher wants to insure that cases with certain characteristics will be included in the sample. First, the population is divided into groups (strata) with the desired features, and then a random sample of cases is selected from within each group (strata). For example, we may want to insure that there are an

---

[6] Imprecise estimates - The sample estimate is not close to the population parameter.

[7] Because the sampling method was not random, the selected farmers do not accurately represent the population.

equal number of rainfed, canal irrigated and pump irrigated farmers in our sample. We define these groups as the strata and then randomly sample for the desired number of cases in each sub-population. The major advantage of stratified random sampling is that it is easy to carry out and it guarantees that there will be enough cases in each major group (strata) that statistically reliable comparisons can be made. In other words, it increases the precision of statistical estimates by maximizing the variability between strata and minimizing the variability within strata (Lipton, 1973 in Norman, 1976). On the other hand, the researcher must have some prior information about the population and sub-populations before the strata can be identified. That is, we must know how the population can be meaningfully subdivided in order to identify strata.

5) Quota-sampling may be employed where a stratified sampling is desired, but lists of the populations units in each strata are unavailable. First, the interviewer randomly draws the sample from the frame. Next, he visits the selected respondent and asks screening questions to determine to which stratum he should be assigned. For example, suppose we want 30 farmers in each of the three strata defined in terms of how the farmers prepares his land. We might ask "Do you plow your field with only human power, an animal, or a tractor?" If the farmer uses a tractor and that strata has not been filled (i.e. less than 30 tractor farmers have been interviewed), we select him as a respondent and complete the interview. On the other hand, if the quota for that strata has been filled, the interviewer does not interview the respondent, but proceeds to the next farmer selected from the population list. This procedure is repeated until the quota for each stratum has been filled (CYMMIT, 1978). The researcher must be cautious when generalizing to the population from data obtained using a stratified or quota sampling procedure. While the farmers may have been chosen at random, the number of units in each stratum is specified by the research design. Therefore, it is only possible to generalize about the population if we know the relative importance of each strata. For example, assume we have sampled in a village from two stratum (rainfed and irrigated farms) — each with 25 respondents. Assume we interview the farmers and find that yields average 2000 and 5000 kg/ha on rainfed and irrigated farms, respectively. We can not estimate average village yields unless we know the relative area in each stratum (weighting factor). Assuming 80 percent of

the area is rainfed, then average area yield can be computed as

$$[(5000 \times .20) + (2000 \times .80)] = 2,600 \text{ kg/ha}$$

In this example, had the researcher failed to weight the data, he would have estimated average yield to be 3500 kg/ha. Hence, if the researcher employs quota or stratified sampling and fails to weight the data, it is impossible to use the results to make accurate inferences about the population.

6) Cluster sampling is a method for selecting distinct heterogenous groups. The clusters are selected at random or systematically from the whole population. For example, the study area may include ten provinces, but it would be too expensive to sample from each location. Hence, three provinces are randomly selected, each constituting a cluster. Sample farmers in each cluster may be chosen by a simple random stratified or quota procedure. The main advantage of this approach is that initially you only need to list tha population of clusters. Only after the sample clusters are chosen, must you obtain a list of the population of farmers. Also, since all of the final sampling units will be located near each other, the cost of data collection will be reduced. The major disadvantage of this procedure is that sampling units in geographically close areas may be more similar to each other than those in more dispersed areas. Consequently, since the sample may be biased the researcher must be cautious in using the sample data to generalize about the population of clusters.

7) Multi-stage sampling combines several of the procedures already mentioned. After the population is initially sampled, the resulting sub-population is again sampled. This procedure can be repeated as many times as desired to build up a survey sample with the required characteristics. For example, suppose we want to draw a sample of farmers in Central Luzon, Philippines, but one-third must be rainfed, canal irrigated, and pump irrigated. In order to reduce travel costs, we list all of the provinces and randomly select three provinces. In each of these three areas we list the population of villages and then randomly select two. In each village, we obtain a list of farmers in each irrigation class. Then, we randomly select a given number of cases from each of the three sub-population. If we desired to have 20 cases in each of the three irrigation classes in each of the two villages in each of the three provinces, a total sample size of 360

cases would be required. The major advantage of this method is that because the initial population (province in this case) is a large unit, it is usually easy to obtain the population listing. In order to select the successive sample, it is only necessary to have a population list for those units selected at the previous level. Also, since the final sample of cases will be located close to each other, transportation costs will be minimized. On the other hand, the main disadvantage is the complexity of the procedure and difficulty in generalizing to the population of clusters from sample estimates.

C. Choosing Between Sampling Methods

Choice of the appropriate sampling method depends on the research problem, available population lists, and financial resources. Common sense provides some guidelines for choosing between the alternative sampling designs for a given research problem. For example, if we wanted to estimate average values and the distribution of those values for farmers in one small village, simple random sampling of farmers would be adequated. If interested in obtaining estimates of these values for a whole province, cluster sampling would be required. Suppose we wanted to compare these values for farmers using modern varieties with those using traditional types in a village where only a few farmers used modern varieties. Stratified sampling would be required to insure our sample had a sufficiently large number of modern variety farmers to conduct the required statistical tests.

In situations where population lists do not exist or are believed to be extremely inaccurate, and we desire to stratify our population in order to make inter-group comparisons, stratified or quota sampling may be appropriate.

When financial resources are limited and the research goal is simply to generate hypotheses, non-random sampling may be adequate. If scientific sampling over a large area is required, cluster sampling will minimize travel time.

As a basic guideline, we should use the simplest sampling procedure that will permit achievement of the research goals. Generally, the simpler the method, the less time and money will be required for implementation. At the same time, we must be aware of the limitations of each sampling method. It is a false economy to use a procedure which is simple to

implement, but does not permit the researcher to make statistically valid generalizations about the target population from the data which is collected.

D. Sample Size

While statistical formulas exist which enable the researcher to estimate the sample size required to obtain a sample estimate of the population parameter[8] at a given degree of precision, such guides are often rather academic. Usually, we are interested in not one but many parameters — each requiring a different sample size to obtain the same level of precision. Furthermore, data on the expected variance of the variables of interest is seldom available prior to doing the survey, so the existing formulas for identifying the sample size for a given level of precision are of little use. As a rough guide, the conventional wisdom is that we need a minimum of 20-25 cases in each of the major sample categories in order to make comparison between these groups.[9] For example, to compare mean yields of rainfed to irrigated farms, we would need 20-25 cases in each strata.

E. Sampling Frames — the Hidden Biases

Once the focus of the survey has been finalized and a basic sampling strategy selected, the researcher must attempt to carry out the sampling plan. This means all available sampling frames (population lists) must be evaluated to determine which is most appropriate. At the village level, several possibilities may exist. The researcher must evaluate each, attempting to determine its inadequacies. Several frequently used lists and their deficiencies are discussed below:

1) Extension agent lists may include only farmers visited by the agent. Use of this frame will probably produce data that suggest an unrealistically progressive, well informed, and successful farm population.

2) Irrigation parcel lists usually include an entry for each parcel in the system.. Thus, farmers with large fragmented holdings would have a greater chance of being selected than individuals farming only one parcel.

---

[8] Population parameter - The numerical characteristic of a population which we try to estimate through sampling

[9] When time and financial resources are not limiting factors and precision is considered to be critical to the research goals, a census of households may be made and the information gained used to estimate sample size.

3) Land tax lists often include only land owners. Consequently, leaseholder and share tenant farmers would be excluded.

4) Head tax lists may exclude women, non-resident farmers and may be incomplete.

5) Census lists may exclude individuals who have arrived since the last census and non-resident farmers.

6) Voter lists are often out of date and may not include non-resident farmers, farmers who don't own land (when property ownership is a condition for voting), and women in areas where only men are permitted to vote.

7) Farmer lists are often out of date and may exclude non-resident farmers.

8) Government production program participant lists will exclude farmers who aren't program participants.

9) Government pump and tractor owner lists may exclude both farmers who obtained the unit outside the official loan program and equipment renters; and will include farmers who no longer own or use the equipment.

10) Maps or aerial photos can provide useful frames when the sampling unit is fields or land. Once the population is defined in terms of a given geographical area, lines are drawn parallel to the horizontal and vertical axes, and numbered consecutively Next, the sample sites are chosen by randomly selecting a horizontal and vertical line and locating the point of intersection. The parcel or farmer identified by this method becomes the sampling unit. This procedure is repeated until the required number of samples are selected. Because large fields have a greater chance of selection, this procedure will over represent large farmers (CYMMIT, 1978).

11) Census. If resources are unlimited, a high level of sampling precision required, and reliable population lists are non-existent, the researcher may choose to conduct a complete census. A short questionnaire is prepared to obtain information about certain variables to be used for stratification — for example, farm size, existence of irrigation, and tractor ownership. Each household (sampling unit) in the village (population) is visited and the desired information collected. If the village is large and a complete census would be too time consuming, a block census may be used. In this procedure, the administrative sub-units of the village are listed and a random sample of these are chosen. Then, the researcher proceeds to

completely enumerate all of the households within the randomly selected sub-units. After completing the census, it is summarized to show the mumber of household falling into each stratification cell. The sample to be interviewed is then randomly selected from the households in each strata. Usually, an equal number of households are chosen to represent each cell in order to guarantee that there are enough cases in each strata to complete statistical analysis.

If households are not dispersed randomly (i.e. all large farmers live in one section and poor farmers in another), the resulting sample will be biased. Also, conducting a census or block census is extremely time consuming and as a consequence, quite expensive. In general, quota sampling accomplishes the same purpose. Hence, a census is most applicable when no sample frame is available and the researcher is interested in determining the sample size, based on knowledge of the variance of a few specific variables. In addition, if data from a stratified sample is to be used to generalize about the population, the census can be used to determine the required weighting factors.

It is unlikely that you will ever find a "perfectly complete" sampling frame at the primary sampling unit level. What is most important is that the researcher be aware of the biases which characterize his sampling frame, before using it — not after the study has been completed. It may be that, given the purpose of his study, the existing deficiencies are unimportant and the available frame can be used. On the other hand, once the biases have been identified, it may be possible to make adjustments to compensate for these defects. In the final analysis, selecting a sampling frame is a compromise. The researcher must weigh the magnitude of the biases that will be introduced by using the readily available frames against the time and money required to obtain a better population list.

V. Primary Data Collection Instruments

Various approaches may be used to collect information required for economic analysis. Each strategy has a particular place in the researchers' toolkit. Yet, it is important to correctly match the data collection instrument with the type of data required.[10]

---

[10] Self administered surveys are not discussed since they have limited use in the rural farm setting in the third world.

Researchers are increasingly recognizing that the standard recall survey can not be used to obtain all of the information we might wish to know, since the accuracy of this data depends on the farmer's ability to accurately remember a great deal of detail. It is useful to make a distinction between "single point" and "continuous" data and between "registered" and nonregistered" data (Collinson, 1972 and Lipton and Moore, 1972 in Norman, 1978). Single point data refers to activities which occur at one point in time while continuous data describes activities which occur over an extended period of time. The continuum ranging from registered to nonregistered data refers to the extent to which circumstances influence the respondent's ability to remember the quantities of an activity". Single point registered data can be collected through a single recall survey long after the event occurred with little loss in accuracy. Examples of such data may include wages paid to hired rice transplanters, and total output from the rice harvest. On the other hand, data which can be characterized as continuous nonregistered must be collected soon after the event — through frequent surveys or farm record keeping — in order to obtain accurate values. Examples of this type of data might include family weeding labor, time spent managing irrigation water, and the harvesting of green corn for household use. This presents the researcher with a dilemma. Sampling error can only be reduced by increasing the sample size or by selecting a homogenous sample. On the other hand, measurement error[11] can only be reduced through frequent interviewing — a procedure which is expensive and therefore necessitates selecting a small sample of respondents.[12] Norman's solution to this problem is to use a large survey to collect single point registered data, but select a sub-sample of these respondents for collecting continuous nonregistered data.

Survey research typically requires that several data collection strategies be used in sequence or combination including the village leader interview, group interview, individual interview, farm record keeping, and direct field measurement. These approaches are described below, along with some advantages and disadvantages of each.

---

[11] Measurement error - The difference between the number or answer given by a farmer and the true correct answer.

[12] Respondent - The individual being interviewed.

A. Village Leader Interview

Typically, survey research is conducted in areas where the researcher has little previous experience. The village leader interview provides an opportunity to orient the researcher to the area. The researcher meets with the head of the village and administers on unstructured[13] questionnaire. While a general set of questions are used to guide the discussion, the purpose of this interview is to stimulate the leader to elaborate on the topics of interest. During this visit the researcher can also collect available secondary data which is recorded by the village leader's staff.

Questions related to aggregate or overall characteristics of the village are particularly appropriate in this situation, since the village leader's responsibilities force him to think in such terms. In addition, the leader typically has a broader world view than the farmers in his village. Hence, it is possible to ask him abstract questions which would be difficult for the less cosmopolitan farmer to answer. Even when an individual farm survey is planned, this initial contact with the leader will enable you to better explain the purpose of your research and facilitate future cooperation. Also, the information gained will enable the researcher to design an individual farmer survey questionnaire which focuses more sharply on the relevant research issues for that village.

On the other hand, village leaders are official representatives of the government. Often, they give biased answers to questions which they may interpret as evaluating the efficiency of their administration. Even the answer to relatively innocent question like "What is the average rice yield in this village?" may be upwardly biased in order to give the impression that his village is more progressive than neighboring areas. Also, overall answers provided by the village leader — although correct — may fail to identify variability in important characteristics of the farming situation. In addition, unstructured interviews must be conducted by skilled enumerators who understand the purpose of the research and can therefore redirect and guide the questioning to obtain the required insight. Finally, since each respondent will be asked slightly different questions, it is difficult to compare the answers given by each respondent.

[13] Unstructured questionnaire lists a number of topics to be covered, but allows the interviewer to modify the order and/or ask additional questions in response to the respondents answer.

The village leader interview foremat can also be easily modified to obtain preliminary information from other community members such as input dealers, product retailer and wholesalers, cooperative officers, and extension agents.

When the researcher is aware of the limitation of the village leader interview, it can be used as an effective tool for rapidly gaining an overall impression of the characteristics of the village and problems facing its farmers. Also, some of the bias characteristic of the village leader interview can be reduced by interviewing two or three leaders separately, and comparing their responses to each question.

B. Group Interview

This procedure also gives the researcher an opportunity to familiarize himself with problems in areas where little previous research has been conducted. The researcher randomly selects five to ten farmers in the village and administers an unstructured questionnaire to the group. Since this procedure is exploratory the interviewers' set of questions should guide the discussion but not restrict it. The farmers should be encouraged to talk openly about the questions raised.

Questions which relate to phenomena which uniformly effect all farmers in the village are particularly appropriate. For example, whether fertilizer is available in the village could be determined from a group interview. When controversial questions are posed, the presence of several persons may facilitate the development of a lively debate which may provide deeper insight into the issue than if the same question was asked to a single individual. Group interviews are also an effective way to develop the appropriate response categories necessary in designing a precoded individual interview (p.19). In response to a question like "What types of off-farm employment opportunities are available to farmers in this village?", the farmers should be encouraged to note all of the existing possibilities.

Group interviews must be skillfully orchestrated to be effective. In some cultures, it is considered impolite to publically disagree. Also, young farmers may hesitate to contradict their elders. Hence, efforts must be made to avoid situations where someone loses face. Questions which might reflect on the effectiveness of the village leader may also receive biased answers since the critic knows he may be censured at a future date. In addition,

group interviews are not appropriate for acquiring information which indirectly measures the competence of the farmer, such as their yield. In answering such questions participants may inflate their answers in order to appear superior. On the other hand, if phrased in a general way ("What are the typical rice yields in this village"), indicative information can be successfully obtained. Finally, unstructured group interviews are subject to the weaknesses of unstructured village leader interview such as noncomparability of data collected.

The group interview is a useful tool by which answers to general village specific information can be obtained rapidly. It is superior to the village leader interview, since the greater number of participants insures that when differences of opinion exists, they will more likely be brought to the surface.

C. Structured Individual Interview

Through the information and insights gained by conducting unstructured village leader and/or group interviews, the researcher can identify specific critical issues to be further investigated. The general hypotheses which guided the village leader and group interviews can than be refined and expressed more precisely. Once this has been done, the researcher can proceed to develop a structured questionnaire[14] which is administered to a sample of farmers.

1) The pre-test interview

Once the questionnaire has been developed, it is necessary to pre-test the data instrument. The pre-test is simply an individual survey which is conducted to evaluate both the validity of the questions and the structure of the questionnaire; and to develop the precoded response categories. The number of interviews required to accomplish this depends on the expected diversity in the population. If a stratified sample is to be interviewed (i.e., rainfed and irrigated farmers in one area or rainfed farmers in several different regions) it would be necessary to conduct five to ten interviews in each strata. On the other hand, if a large sample is to be drawn from one relatively homogenous area, a total of as few as ten-fifteen interviews would be adequate.

---

[14] A structured questionnaire includes a specific set of questions that are asked to every respondent in the same order.

2) The farmer interview

After the pre-test has been completed and necessary changes made, the researcher is ready to conduct individual farmer interviews. In this procedures the enumerator isolates the farmer and asks a set of structured questions.

The greatest advantage of the individual interview is that it permits each respondent to express his opinion on each question, without being influenced by others. Also, since the same set of questions are asked to each respondent, response variability will be captured. Once the data has been collected, it will be possible to statistically analyze the farmers' answers and make inferences about the population, based on the sample responses. Finally, the less structured the schedule, the more experience is required to obtain accurate and comparable answers. Hence, a properly designed and pre-tested structured farmer interview can be administered by an inexperienced enumerator.

The individual interview is not without weakness, regardless of its popularity. First, in the village setting it is difficult to really isolate the respondent. Often relatives are present who try to influence the farmer's answers. When a village leader is used to introduce the researcher to the farmer, his continued presence is likely to bias the respondent's answers. Second, interviewers differ in their ability to explain the meaning of complicated questions. Hence, when several interviewers are employed, different respondents may actually be asked slightly different questions for the same item. When this occurs the answers to that item are of little value. Third, it is very time consuming — and therefore costly — to conduct an individual farmer survey. While the interview should never take more than 1-2 hours, a great deal of time may be lost in simply locating the respondent. Hence, it is difficult to complete more than three interviews per day per enumerator. Finally, timeliness is critical for obtaining correct answers to questions which are time specific. The longer the length of time between when the event of interest occurs and when the interview is conducted, the less accurate the answer.

D. Farm Record Keeping

For certain types of economic analysis — such as costs and returns studies or farm level production functions — very accurate detailed

information is required. Particularly when non-point nonregistered data is sought, farm record keeping is required. This approach to data collection may be thought of as a farm survey in which a fixed set of questions are repeated during each meeting with the farmer. When detailed labor and consumption data are sought, the enumerator typically visits the farmer once a week and asks him to recall his activities during the previous seven days. Yet, some individuals will find it difficult to recall information over the previous week. If someone in the household is literate it is often possible to ask the farmer to record the desired information in a record book or on a calendar. Then, when the researchers visit the household, these data can be reviewed by the researcher and permanently entered into a record book.

The most desirable feature of farm record keeping is that the system minimizes the elapsed time from when an activity is completed to when it is recorded. Hence, it is possible to obtain very accurate data such as the timing of input application, quantity of inputs applied, labor use, and household expenditures. Also, since the interviewer meets the farmer frequently, there is ample time to gain the farmer's confidence and clarify any suspicions he might have about the interviewer's motives. In addition, since periodic visits are made, it is possible to clarify any doubts about answers given during a previous meeting.

The major disadvantage of the farm record keeping approach is that it takes a full season or more to obtain the desired information. Also, unless the sample size is small, it will be quite costly to employ a staff of field enumerators for a complete cropping season. Often the cost of farm record keeping can be greatly reduced by employing part-time village enumerators. School teachers who live in the village are particularly appropriate to assist the researcher because they are generally responsible, literate, and well respected by the villagers. After a one-day orientation course, each teacher can be assigned 5-10 farmers to interview each week. At each meeting, information about the previous week's activities is recorded. Yet, when this approach is used, the supervisor of the project must periodically examine each record book to insure that accurate data is being collected and check with the sample farmers to guarantee that the village enumerator is actually visiting the farmers weekly.

E. Verification Measurements

Surveys generally rely solely on the respondent to supply all of the required information. Yet, it is often quite easy to increase the accuracy of the information collected by making simple measurements to verify the farmer's answer.

1) Physical measurements

If the researcher suspects that the farmer does not know the size of his parcel, it is desirable to hire an engineer to survey the farmer's parcels.

If the survey is conducted before harvest time, it may be possible to visit the farmers parcels and randomly select and mark off a portion of the field to be crop cut.[15] Then, the farmer is requested to harvest these areas and measure the output from each plot. When the researcher returns after the harvest, these results can be recorded. While asking the farmer to do the crop cut by himself may introduce measurement error, it is extremely time consuming for the researcher to supervise this activity. Also, it is almost impossible to determine beforehand the actual day the crop will be harvested unless the field staff actually stays in the village throughout the harvest period. Consequently, when the sample size is large, the self-administered crop cut is the only practical strategy for obtaining an objective yield estimate.[16]

Researchers frequently find that farmers report their output, using several different units of measures. Meaningful yield comparisons can only be made if all of these measures can be converted to a common unit. For example, Indonesian farmers typically report their rice harvest as bundles of stalk paddy, paddy (threshed, but not milled), and beras (milled paddy). Before beginning the survey, the researcher must decide which form will be the standard. If all output is to be converted to paddy, the researcher must ask the farmer how many kilos of paddy is normally obtained when one

---

[15] See Hunt, R.E., 1969 for procedures for conducting crop cuts.

[16] Recent studies suggest that crop cuts may give upwardly biased yield estimates.

bundle of stalk paddy is threshed. If the farmer can not answer this question, the researcher should randomly select three or four bundles, thresh each, and weigh the grain in order to estimate the appropriate conversion factor. Similar procedures should be followed to convert the output of other grains and legumes to shelled equivalent and root crops to wet tubors weight. When output is expressed by the farmer in volume measures such as tins of grain, the researcher must either ask the farmer the weight of one such unit or — if possible — actually weigh the grain contained in the tin.

2) Time measurements

In some research studies it is necessary to obtain accurate estimates of the time required to complete specific tasks. To obtain this data, the enumerator can go to the field with the farmer and measure with a watch the time actually spent on each field operations. To reduce the time required to obtain this data, the enumerator may choose to measure only the time spent to complete part of the job — such as harvesting one row of soybean. This data is then used to compute the labor input require to complete the task on one hectare of land.

This approach is only appropriate when the greatest degree of accuracy is required in developing estimates of time spent on specific operations. It is particularly useful if the researcher believes that the labor input reported by the farmer includes a large component of slack time which inflates the labor input estimates. Yet, this approach does not insure unbiased data. Aware that he is being timed, the farmer may work faster than normal while the researcher is observing his activities and not take the normal rest breaks.

3) Farmers' field experiments

Survey data often fails to show the potential of a given technology because farmers do not use it, apply it incorrectly, or only apply levels specified in government production programs. One method to estimate the yield potential of a given package of practices is to actually conduct simple experiments in farmer's fields. This type of activity requires a great deal of technical supervision and is beyond the range of typical

survey research operations. Yet, it is a particularly appropriate strategy for integrating survey and agronomic research that is directed at evaluating constraints to higher farm yields.[17]

F. Choosing the Appropriate Primary Data Collection Instrument

Several types of primary data collection instruments are available to the survey researcher. Each is appropriate for obtaining a particular kind of data which enables the researcher to carry out specific research goals. Consequently, the objectives of the research project must be explicitly and clearly outlined in detail before the researcher can decide which type of survey is required.

In situations where only general information is required, the village leader or group interview may be adequate. On the other hand, if an individual interview is necessary, it will be almost impossible to construct a focused questionnaire unless a village leader or group interview is first used to collect preliminary data which can be used to refine general hypotheses into specific hypotheses. In addition, it will be absolutely necessary to pre-test the draft questionnaire. If continuous nonregistered data is needed to achieve the research goals, farm record keeping will be required. In the same vein, the verification measurements are most applicable to surveys and farm record research which place a high value on accurate measures of area, output, labor input and technology potential.

Often the research goals can be met through the use of a single research strategy. On the other hand, it may be appropriate to use a combination of or all of these strategies, each designed to generate data for a specific component of the overall research effort.

VI. Survey Questionnaire Development

Because most farm level research relies upon the individual interview, this section focuses in somewhat greater detail on that approach. While survey data will always be subject to some degree of inaccuracy, several potential sources of error can be avoided by taking appropriate precautionary measures in developing the questionnaire.

---

[17] See DeDatta, S.K. et.al. 1978.

To design a good questionnaire, the researcher must first have a clear ideal of all hypotheses to be tested. Only after clarifying the purpose of the survey can questions be developed to solicit data appropriate for testing each specific hypothesis. In addition, the researcher must have a good understanding of the farming situation in the area where the research is to be carried out. The village leader interview is particularly appropriate for obtaining such location specific information.

A. Questionnaire Design

It is an extremely difficult task to design a good questionnaire. Consequently, the initial draft should be completed under the supervision of the primary researcher, since only this individual has a comprehensive understanding of the objectives of the survey. Also, the individuals designated to analyze the data must be actively involved at this formative design stage to reduces the likelihood that essential questions — necessary for the testing of specific hypotheses — are left out of the questionnaire.

1) Organization of topic sections

The first section of a questionnaire is the cover page. Here the respondent's name, village, section in village, and similar information is recorded which would assist in locating the individual if it were necessary to make a return visit. Usually each respondent is given a code number which refers to the sequence of the interview (i.e. 1 = first person interviewed, etc.). Since pages in the questionnaire may become separated from each other, it is advisable to copy this code number onto the top right hand corner of each page. The cover page also should include information about the interview such as the name of the enumerator, starting and ending time, date conducted, name of the person who checked the questionnaire and the date it was reviewed.

The subject matter to be covered during the interview must be grouped into sections which focus on a single topic, such as land tenure, marketing, and input use. If the questions are not organized into homogenous sections, the respondent will have difficulty in providing the requested information and the pace of the interview will be slowed.

The sequence in which the sections are ordered is also important. Each must be placed so as to facilitate the introduction of succeeding sections.

If wisely ordered, answers to questions asked early in the interview will provide the enumerator with the background needed to judge if the answers to later questions are complete and consistent. For example, an inventory of land "owned and/or operated" should precede specific questions about inputs applied to each parcel. The identification of all land resources will help guarantee that input information is obtained for all parcels. Similarly, questions on household members' non-farm employment should be preceded by a listing of household members.

Often an interview schedule will include questions that can best be answered by someone other than the respondent. For example, the farmer's wife may have a more accurate idea of family food consumption than the farmer. To facilitate interviewing, it is best to include a separate section for each family member who will be specifically interviewed.

The sections included at the beginning of the interview should help establish rapport with the farmer. Hence, they should be relatively easy to answer and focus on non-threatening information. Since individuals do not like to reveal information on such topics as income, these questions should be placed in the latter portion of the schedule, if asked at all.

It is also good to alternate simple-to-answer sections with more difficult ones. For example, a section of questions which require the farmer to recall detailed information on input use for the past season might be followed by a section which only requires a yes or no answer.

2) Ordering questions in each section

Every time the respondent is required to move from one section to the next, it is necessary to first describe the type of questions which are to follow. New sections should be introduced by a short statement like "This section of questions will focus on the type of machinery you owned or used during the 1977-78 wet season".

It is easier for people to recall more recent occurances than events which happened in the more distant past. By asking about recent events first, an opportunity is provided for the respondent to first became familiar with the meaning of the question before being required to recall more obscure data. This strategy is applicable to such questions as output history, previous employment, and crop sales. When asking about events that occured

in the distant past, more accurate data can be obtained if you associate the year with an event that occured at that time. For example, "What varieties did you plant in 1973 (the year brown planthopper destroyed a lot of rice)?

Care must be taken to avoid biasing the farmers answers to a question by items asked earlier in the section. Suppose we first ask "What are the names of the most important insect pests which reduced your rice crop yield during the 1977-78 wet season?" If this question is followed by "What were the main biological or agronomic causes of yield loss to your rice crop during the 1977-78 wet season?", respondents would probably answer, insects. Such biased responses can best be avoided by reversing the order of the two questions. Yet, care should be taken to word general questions as precisely as possible in order to obtain the desired type of response. If the word "biological or agronomic" had been left out of the above question, some farmers would have responded with answers such as "fertilizer is too expensive" and "don't own a sprayer to apply insecticide".

It is also a good practice to open a section with requests for more familiar information, followed by questions for less familiar data. For example, if the researcher is interested in information about farm prices for paddy in various markets, he would order the items as follows:

"What was the price/kg of paddy at harvest time (April 1978) sold at the farm gate?" _____

"What was the price/kg of paddy at harvest time (April 1978) sold in the village market?" _____

"What was the price/kg of paddy at harvest time (April 1978) sold in the district market?" _____

The first question is more familiar since most farmers will know the farm gate price. The last question is least familiar since the farmer may not sell rice in the district market.

B. Wording of Individual Questions[18]

All questions must be carefully worded in order to clearly communicate to the respondent the exact meaning of the inquiry and type of response

[18] Some of the examples in this section illustrate problems identified by Frank Lynch, undated, and Prewitt, 1975.

expected. Unless care is taken in designing each question, the data collected will be of little use. The following example illustrates how poor wording can be responsible for the incorrect reporting of data. Suppose we are interested in learning about amount of rice produced on all land farmed by household members. Assume the household farms one parcel. From their 1.0 ha plot they harvested 5.0 tons of paddy in the 1977-78 wet season and 1.0 ton in the 1978 dry season. We might ask:

"How much rice did your family harvest this past year?"

Several farmers with the same actual output could give very different answers as shown below:

Farmer	Answer	Farmer's interpretation
1	6.0 tons	correct
2	5.0 tons	Farmer only reported the main crop which was grown in the wet season, because the dry season harvest was so low.
3	3.0 tons	Farmer was a share tenant. He gave half of the crop to the landlord, so he reports only what he received.
4	4.8 tons	Farmer was an owner operator, but gave 1/5 of the crop to the laborers who assisted with the harvest. He did not consider this amount part of his family's share.
5	5.0 tons	Farmer's son managed the 1978 dry season crop and used the output to pay for school fees. Hence, the farmer did not include this as family income.
6	4.0 tons	Farmer sold all of the harvest immediately after harvest as milled rice. Hence, the output figure reported is milled rice equivalent and not paddy. (1 ton of paddy = 0.68 tons of milled rice).

We see that because the question was not clearly worded with each term explicitly defined, six different farmers gave entirely reasonable -- but different -- answers. If the researcher were to accept each answer as correct, this estimate of the sample's output of paddy would be extremely inaccurate. Yet, all of these problems could have been avoided if the question had written as follows:

During the 1977-78 wet season, how many tons of paddy were harvested from each of the parcels operated by household members?

Parcel 1: Harvester's share _____

Thresher's share _____

Landlord's share _____

Farmer's share _____

TOTAL _____

This format should be repeated for each parcel and each season in order to insure that all required data is reported.

Failure to communicate the precise meaning of a question can be due to several different causes:

1) Multiple questions — Each question should focus on a single point. In general, short questions are less likely to include multiple questions.

Poor wording: "During the 1977-78 wet season, did you operate a farm and engage in non-farm work? ___

Clearer wording: "During the 1977-78 wet season, did you operate a farm?" ___

"During the 1977-78 wet season, did you engage in non-farm work?" ___

CODE: 0 = no; 1 = yes.

2) Vague terms — "Often", "many", "frequently" and similar adjectives do not have the same meaning to everyone and should be avoided in questions designed to obtain quantitative information.

Poor wording: "During the 1977-78 wet season, how often did you spray your rice parcel with insecticide?" ___

Clearer wording: "During the 1977-78 wet season how many times did you apply insecticide on your operated rice parcel ?" ___

3) Technical terms — Agricultural jargon and other technical terms may not be familiar to farmers. Whenever possible, descriptive phrases should be used to explain a technical term.

Poor question: "During the 1978 wet season, did you apply a basal dressing of fertilizer on the rice parcel you operated? ___

Clearer question: "During the 1978 wet season, did you apply fertilizer before plowing on the rice parcel you operated?" ____

CODE: 0 = no; 1 = yes.

4) Overlapping answers — All questions should have only one possible answer.

Poor question: "During the 1977-78 wet season, did you use human labor, a carabou or tractor for land preparation on the rice parcel you operated?"
(Note: Farmer may use a tractor for the first plowing and a carabou for the second plowing). ____

Clearer question: "During the 1977-78 wet season, what source of power did you use to prepare the rice parcel you operated?" ____

CODE:
0 = no rice parcel;
1 = human labor only;
2 = carabou only;
3 = tractor only;
4 = human and carabou;
5 = human and tractor;
6 = tractor and carabou.

5) Time-specific questions — Farmers do not follow the same practices every season. Accurate information can only be obtained with time specific questions.

Poor question: "Did you plant high yielding varieties on the rice parcel you operated?" ____

Clearer question: "During the 1977-78 wet season, did you plant only high yielding varieties on the rice parcel you operated?" ____

CODE: 0 = no; 1 = yes.

6) Degree of adoption — Many practices are only partially adopted by farmers. If accurate information is required, the degree of adoption must be specified.

Poor question: "During the 1978 wet season, did you plant high yielding varieties on the rice parcel you operated? ____

Clearer question: "During the 1978 wet season, how many hectares of your rice parcel was planted to:

high yielding varieties ____

local varieties ____

Total area ____

(NOTE: ENUMERATOR COMPUTES[19] PERCENT OF AREA IN HYV's) ____

7) Continuous phenomena — Many phenomena which effect agricultural production do not occur as discrete events, such as drought and crop failure. Consequently, numeric estimates of the intensity of the event should be sought. While farmers can not be expected to answer these questions precisely, a quantitative guess is more accurate than a yes/no answer.

Poor question: "During the 1977-78 wet season, did your rice crop suffer from drought?" ____

Clearer question: "During the 1977-78 wet season, how many days was there no standing water in your rice field," followed by ____

"Did a shortage of water reduce your rice yield?" ____

CODE: 0 = no; 1 = yes.

8) Unit of observation — The preceding examples assumed that the farmer operated only one parcel of rice land. More commonly a farm will consist of several parcels with the farmer managing each parcel of land in a unique way. Consequently, data on inputs applied and management practices used should be asked with respect to each parcel if precise data is required. Of course, the degree of precision necessary depends on the research goals.

Poor question: "During the 1977-78 wet season, how many days old were the seedlings that you transplanted?" ____

Clearer question: "During the 1977-78 wet season, how many days old were the seedlings you transplanted on each parcel operated?"

Parcel 1 ____

Parcel 2 ____

Parcel 3 ____

---

[19] Instructions to the enumerator are always written in capital letters to distinguish these sentences from interview questions.

Alternatively, general input-output data can be collected on a parcel basis. Then one parcel can be selected for intensive investigation and detailed questions such as age of transplanted seedlings and date of input application can be directed at only this specified "intensive data parcel".

9) Biased association — Farmer's answers to questions can be biased by associating the question with institutions or individuals towards which people have intense feelings, such as the church and political leaders. Valid data can only be expected if the question is neutral.

Poor question: "Do you believe President Marcos's land reform program has helped farmers in this village achieve ownership of land?" ___

Clearer question: "Do you believe land reform has helped farmers in this village achieve ownership of land?" ___

CODE: 0 = no; 1 = yes.

10) Expertise —Questions directed at obtaining factual information assume the respondent is capable of supplying the solicited data. Often it is necessary to use screening questions to first determine if the respondent's answer to the following question is to be trusted. Suppose we are interested in evaluating the timeliness of fertilizer delivery.

Poor question: "During the 1977-78 wet season, was fertilizer available at the local supply store when you wanted to fertilizer your rice land?" ___

Clearer question: "During the 1977-78 wet season, did you grow rice?" ___

"During the 1977-78 wet season, did you apply fertilizer on your rice land?" ___

CODE: 0 = no; 1 = yes.

(NOTE: IF NO TO EITHER OF THE ABOVE, SKIP THE FOLLOWING QUESTION)

"During the 1977-78 wet season, was fertilizer available at the local supply store when you wanted to fertilize your rice crop?"

CODE: 0 = no; 1 = yes.

The first two of these three questions are called "screening questions" and are intended to determine if the respondent is qualified to answer the item that follows. In the above example, it would be foolish to ask the farmer about the timeliness of an input applied on rice he doesn't grow rice or use the input.

11) Aggregated questions — As noted earlier, data about events that occur over time are best obtained through record keeping. Yet, on some occasions we may include such questions in a survey interview. Generally, more accurate information can be obtained if the question is broken down into its basic components. Assume we are interested in estimating total paddy sales from the 1977-78 wet season crop harvested in February.

Poor question: "From your 1977-78 wet season harvest, how many kg of paddy did you sell?" ____

Clearer question: "From your 1977-78 wet season harvest, how many kg of paddy did you sell during the following months?"

March	_____
April	_____
May	_____
etc.	
Total	_____

While farmers will have great difficulty answering accurately a question about previous sales, we will obtain more accurate information by first seeking monthly data that can be summed to obtain a yearly figure than by asking directly for a yearly total.

12) Leading questions — Questions should be worded so that the farmer can not tell what answer is preferred by the enumerator. Suppose we want to determine if a farmer follows a certain traditional practice.

Poor question: "During the 1978 wet season, did you follow the superstitious practice of giving food to the rats so they would not damage your rice crop?" ____

Better question: "During the 1978 wet season, did you give food to the rats so they would not damage your rice crop?" ____

CODE: 0 = no; 1 = yes.

13) Unrealistic questions — A question which asks for information that the farmer is unlikely to remember or simply does not know should not be asked. Once an response is obtained, individuals unfamiliar with the inaccuracy of the answer will use the data as if it were a fact. Examples of information difficult to recall include detail inquiries about money spent on various consumption items such as monthly food, clothing, medical purchase and the dates on which various farming operations were completed. An example of the type of information farmers are unlikely to know is the market value of their land.

C. Format

The time required to complete each interview can be reduced and the accuracy of the data increased if careful attention is paid to the way questions are laid out.

1) Parallel questions — Often several questions are asked which are quite similarly phrased, except for one or two words. The researcher can effectively utilize the famliarity which is achieved through asking the first question to introduce the succeeding questions.

Poor format: "Did you buy any fertilizer during the 1977-78 wet season?" ___

"During this wet season of 1977-78 did you purchase any insecticide?" ___

"Was any rat poison purchased for use on your farm during the 1978 wet season?" ___

Clearer format: "During the 1978 wet season, did you purchase any:

fertilizer ___

insecticide ___

rat poison ___

(REPEAT QUESTION FOR EACH INPUT)

CODE: 0 = no; 1 = yes.

This format not only saves space, but also guarantees that an identical sentence pattern and word order is used for questioning about each input.

2) Answers to open questions -- Some questions may require the interviewer to write in the farmer's response. Care should be taken to make sure that answers can actually be written in the space provided. Also, lines should be provided onto which the answer can be written.

Poor format: "What were the two most important agronomic or biological reasons why your rice yield was not higher?" _____

Better format: "What were the two most important agronomic or biological reasons why your rice yield was not higher?"

1. _____
_____
_____

2. _____
_____
_____

3) Closing open questions -- Questionnaires typically include items which require non-numeric responses.[20] For example, if a farmer is asked "Where did you market your rice?", he might answer one of the following: trader, cooperative or neighbor. If we permit the farmer to answer in his own words, we call this an "open" question. Yet, in order to summarize this type of data, the researcher will have to count the number of responses of each type. A question is "closed" if we only permit the farmer to select one or more of a set of possible answer. By precoding all possible response categories, analysis can be completed much more rapidly. In addition, when the questionnaire is precoded the interviewer can conduct the interview more rapidly, since it will be necessary to only write in the code number, not the words. Finally, respondents often answer open questions in so many different ways that it is impossible to make any sense out of "open" answers. An example of a precoded "closed" question is shown below:

---

[20] Non-numeric responses - Answers to questions like "where do you market your rice?", in contrast to numeric response or answers that are given as numbers such as output and labor input.

"During the 1977-78 wet season, what off-farm jobs did you engage in?"

First job ____

Second job ____

CODE:

0	=	none	3	=	weeding
1	=	plowing	4	=	harvesting
2	=	transplanting	5	=	other

Questions which require a no or yes answer should also be precoded, by assigning the value 0 to a no answer and 1 to a yes response.

4) Answer to closed questions — In order to make data processing easier, the answer to such question should be separated from the question itself. As shown in the preceding sample questions, when numeric or closed answers are required the answer should always be placed on a blank line to the right of the question.

5) Question numbering — The careful numbering of questions also assists in processing the data. Each main section should be given a separate number. Next, each question in that section is given the section number plus a number to indicate its position in the section. Finally, questions with several parts are given a third number to indicate the order of the parts within the question. The following example illustrates this procedure.

1.0 Farmer's Socio-Economic Characteristics

1.1 "How many years old are you?" ____

1.2 "How many children live in your household?" ____

1.2.1 Males under 15 years ____

1.2.2 Females under 15 years ____

2.0 Employment Activities of Farmer

2.1 "During the 1977-78 wet season, did you work for pay on anyone else's farm?" ____

2.2 "During the 1977-78 wet season, did you work for pay in any non-farm activities?" ____

CODE: 0 = no; 1 = yes.

Notice that all section headings (1.0, 2.0), question headings (1.1, 1.2, 2.1, 2.2) and question sub-sections (1.2 1, 1.2.2) are indented the same

number of spaces. This organization makes it easier for the interviewer to administer the questionnaire.

6) Condensed tables — Frequently it is useful to construct a table for recording information that would require several pages to obtain if asked through separate questions. When a table is used, it should be preceded by a set of questions to be asked in order to properly introduce the table, as illustrated below.

LAND OPERATED AND OWNED

We are interested in learning about the type of land you operated and owned during the 1977-78 wet season. How many hectares of land do you own or operate in each of the following categories?

(NOTE: 1.1 REFERS TO THE FIRST RAINFED PARCEL, 1.2 REFERS TO THE SECOND RAINFED PARCEL, ETC.)

Table 1. Hectares of land by tenure class and land type

	Land Operated			Land owned but not operated	
	Owned	Rented-in	Shared-in	Rented-out	Shared-out
Rainfed 1.1	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __
1.2	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __
Irrigated 2.1	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __
One crop- 2.2 ping season	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __
Irrigated 3.1	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __
Two crop- 3.2 ping season or more	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __
Total	__ __ . __	__ __ . __	__ __ . __	__ __ . __	__ __ . __

DEFINITIONS:

- Parcel - a separate piece of land of one tenure and land type.
- Rented - a fixed payment is paid for using the land.
- Shared - a portion of the harvest is paid for using the land.
- Rainfed - the only source of water is rainfall.

Several pieces of information are included to guide the correct administration of the questions. First, an introduction statement about the purpose of the section is followed by the specific questions. Second, the "Note" explains the meaning of the numbers associated with each row entry in the table. Notice that these instructions are written in capital letters so the interviewer can easily separate the information that is intended to explain the question from the actual question. Third, the "definitions" explain the exact meaning of key words. This information should be included on the same page as the questions to facilitate quick reference.

7) Direct Key-punching of data -- Often, researchers are advised to design the questionnaire in order that the data collected may be directly key-punched onto IBM computer cards. Yet, few researchers in developing countries have access to computers. Also, computer time is extremely expensive and labor is relative inexpensive. Consequently, in most situations, it is not advisable to prepare the questionnaire for direct key-punching. On the other hand, once the data has been tabulated, the researcher may decide that certain hypotheses can only be tested using multivariate analysis. At this time, a subset of the variables can be transfered to IBM coding sheets, key-punched, and subjected to computer analysis.

In the rare situation where no preliminary analysis is planned and all data will be punched for computer analysis, it is possible to design the questionnaire so as to simplify data coding. The most effective way to streamline coding is to develop a pre-numbered questionnaire (see Appendix 2). In other words, when the schedule is designed, a number — corresponding to a column on an IBM computer card — is assigned to each blank space into which the farmer's answer is entered.

The increase in survey efficiency achieved through pre-numbering is due to the fact that interviewing and computer coding are merged into a single step. If properly pre-numbered, the data can be read directly by the key-puncher -- eliminating the need to transfer data to code sheets before key-punching.

Successful use of a pre-numbered questionnaire depends on the degree to which the researcher can anticipate the farmers' answers. If direct punching of the data is planned, all open questions must be converted to closed questions. Also, it is advisable to leave at least one extra space for numeric responses in anticipation of values larger than those encountered

in the pre-test. Finally leave at least two space for non-numeric answers in order to enter additional response categories for unanticipated answers to closed questions.

8) Questionnaire length — The collection of accurate data requires that the farmer pay careful attention to each question asked. Every experienced enumerator knows that the longer the interview, the greater the difficulty in holding the farmer's attention. An hour or less is probably the ideal length of time for an interview. If it will take more than two hours to complete the questionnaire, it is best to conduct the interview in two sessions — each on separate days. In addition to reducing respondent fatigue, splitting the interviewer into two sessions gives the supervisor an opportunity to edit the first set of questions before the second visit. Then, during the second visit, the enumerator can ask again any questions which were omitted or otherwise improperly completed.

D. Operations Manual

Most questionnaires include a large number of questions covering a wide range of topics. Many of these questions include words which have very specific meanings. It is unrealistic to expect the interviewer to remember the correct way to ask each question and the meaning of each keyword. This is particularly true when young inexperienced enumerators are employed to conduct the survey. Consequently, it is advisable to develop an operations manual which can be referred to by the enumerator for guidelines when unsure of a particular point. These instructions should supplement the information and definitions included on the questionnaire itself. A set of instructions should be provided for each section. The following paragraph illustrates the type of information that is needed to explain the section on "Land Operated and Owned" (p. 37)

Table 1 is a record of all parcels of land operated or owned but not operated by the respondent during the 1977-78 wet season. A parcel is a contiguous (all together) piece of land with a single type of irrigation and single tenure type. For example, a piece of land that is half rainfed and half irrigated must be entered as two parcels. Enter the area (ha) of each parcel of each class of land in the table under the appropriate tenure column.

The numbers in the far left column are the parcel codes. For each two digit number, the figure to the left of the decimal refers to the land class (i.e. 1 = rainfed, 2 = irrigated, 1 rice crop per year, 3 = irrigated, 2 rice crops per year). The number to the right of the decimal denotes the first (1), second (2) etc. parcel of that type. For example, a hectarage entered on row 3.2 would indicate that this was the farmers' second irrigated parcel where two crops or more are grown each year. If a respondent has more than one parcel of a given type, enter the one with the largest area (ha) first.

E. Translating

If the principal researcher is not fluent in the language spoken by respondents, the draft questionnaire must be translated into the target vernacular. This is a particularly critical stage in survey development, since translation errors frequently distort the meaning of questions and as a consequence the respondents' answers do not address the intended research issue. It is naive to assume that translating is a purely mechanical activity that can be successfully undertaken by anyone who knows the target language. A good translater should not only be a native speaker of the target language, but also fluent in the principal investigator's native language, experienced in survey research, and knowledgeable about the activities at which the questionnaire is directed. For example, if rice farmers are to be interviewed, the translator should be knowledgeable about rice farming.

There are four steps which must be followed in order to insure that the translation is correct. First, once a translator is selected, the principal investigator should spend considerable time explaining the purpose of each question. Second, once the translator fully understands the purpose and content of the survey instrument, he can proceed to translate each question into the vernacular. Third, the translated version of the questionnaire is given to a second native speaker to translate back into the primary investigator's native language. Then, the wording of each of these questions is compared with the original to determine if the meaning of each item has been preserved. Double translation is essential for reducing translation error. Through this technique, it will be possible to identify words and grammatical constructions which are particularly difficult to translate and errors arising from misunderstandings which the principal translator might have had as to the intent of specific questions. Finally, after identifying the problem words or questions, the two native speakers and the primary investigator can get together and resolve apparent discrepancies.

Particular care must be taken in translating grammatical constructions, concepts, and words which do not exist in both the original and target language. For example, in Tagalog there is a direct translation for white-head (uban), the characteristic sign of stemborer damage. On the other hand, there is no widely recognized popular term for brown planthopper or green leafhopper. In the same vein, cultures differ in the way in which the color spectrum is divided. In Mende (a West African language), no distinction

is normally made between yellow and orange. Hence, both colors are represented by a single word. In such instances where translation is paricularly difficult, it is best to consult a linguist in order to identify the most appropriate manner by which the concept may be accurately conveyed.

If the target farmers do not all speak a single language, the questionnaire must be translated into all relevant languages. The field enumerator should never be allowed to translate the questionnaire into the local vernacular while conducting the interview. Permitting such a practice will greatly reduce the accuracy of the data collected, since the untrained translator is likely to unintentionally change the meaning of sensitive questions.

VII. Pre-testing the Draft Questionnaire

Once the questionnaire is drafted, it must be pre-tested. The pre-test is necessary to determine if the enumerators are likely to encounter any problems in administering the questionnaire.

A. Personnel

Conducting a pre-test is a difficult task. Only experienced enumerators are capable of observing problems that the respondent might have in answering the questions. Consequently, only the research team leader and one or two senior interviewers should participate in this activity.

B. Location

To be effective, farmers typical of the target population must be interviewed. Three points should be considered. First, pre-tests must be made in all target agro-climatic areas where the survey will be conducted. For example, a pre-test in an irrigated rice area will not uncover problems that may arise if the survey is carried out in a rainfed or upland environment where mixed cropping is common. Second, the farmers interviewed must have the same socio-economic characteristics as the target population. Interviewing only large farmers will not provide reliable information as to the questionnaire's appropriateness for surveying small farmers. Third, field testing in or near the target area is necessary to identify translation problems which arise due to differences in the dialect spoken by the translator and the target population, existence of local idioms, and similar problems.

C. Indicators of Problem Questions

During the pre-test, the researchers should look for any difficulties the farmers have in understanding the question which can arise from poor translation, unfamiliar terms, complicated grammatical constructions, and illogical ordering of questions. Indications of questions which need to be carefully evaluated include

- the same answer given by each respondent
- questions which the respondent could not answer
- answers that indicate the farmer did not understand the question
- questions which took the respondent considerable time to answer.

Often, simply rephrasing the question will remedy the problem. For example, it may be impossible for a farmer with a limited world view to answer the question "If you were the Minister of Agriculture, what would you do to improve agriculture in this village". Yet, simply rephrasing the item to read "What could the extension service do to help you increase yields?" could be more easily answered.

D. Indications of Format Problems

The questionnaire must also be evaluated to determine if the interview schedule is structured so that the researcher can fill in the answers without a great deal of difficulty. The key to sorving this problem is to word the questions the way farmers think about the item. For example, researchers often ask for production data in terms of yield/hectare. Because farmers seldom think in such terms, the interviewer must take the farmers estimate of output, convert the local land unit to hectares and then divide this value into into total output. Not only is it possible that the researcher will make an error in calculating yield, but also this effort will consume a great deal of time and as a result, prolong the interview. Consequently, it is better to ask the farmer the output in the local unit of measure and record this response. Then, the interviewer can compute yield/hectare after returning to the office.

Numerous other problems like this are likely to be identified when pretesting the first draft of the questionnaire. One indication of a poorly constructed question is notes or calculations in the page margins. When the interviewer can not directly enter the farmer's response into the space provided in the questionnaire the format must be modified to insure accurate and rapid recording of the farmers' answers.

E. Inadequate Precoding

Success in using a precoded format depends on the researcher's ab to identify all of the probable answers before the survey questionnaire finalized. While it is always possible to include the category "other" a large number of responses falls into this group the value of the data collected will be reduced. Consequently, when constructing a closed qu include a blank for an open response if the closed alternatives are not applicable, as shown below. Then, after the survey is completed the "o responses can be recoded into specific categories.

During the 1978 wet season, whom did you consult before choosing the variety you planted?

CODE:					
	1	=	no one	5	= village leader
	2	=	wife	6	= input dealer
	3	=	extension agent	7	= other _____
	4	=	other farmers		

VIII. Selecting Enumerators

Regardless of care taken to develop a good survey schedule, the rel lity of the data ultimately depends on the person interviewing the farme Unfortunately, the selection and training of enumerators is usually over looked in the rush to get into the field and begin the survey.

A. Characteristics of Good Enumerators

Reseachers have found that the quality of the interview is strongly influenced by the personal characteristics of the enumerator. Several p need to be taken into consideration (Norman, 1977, CYMMIT, 1977, Herdt a Bernsten, 1975).

Personality - friendly, interested in rural problems, respects loca traditions, willing to live at the research site.

Language - fluent in both the local dialect and the language (id spoken by the farmers, can express himself clearly.

Education - literate, some knowledge of the subject matter (local farming methods for a farm survey) that is the focus the survey, preferably with high school education.

Experience - lived in a community similar to where the survey will conducted, previous survey activities. Also, individu with a sales experience often are particularly good a interviewing.

Sex - depends on the culture and subject of survey. In the Philippines, farmers generally prefer to talk to women interviewers. In Moslem areas, men often make better interviewers. If interviewing women about food consumption, female interviewers may be more effective.

Age - older interviewers are generally more effective than younger individuals.

B. People Who Make Good Interviewers

Various types of individuals may be employed who meet the above criteria.

Agricultural research staff who know the local language are excellent interviewers. Arrangements should be made to insure that they will be available for the duration of the survey. Since these individuals often have other responsibilities, this is often difficult. Yet, it is impossible to manage a survey effectively if several interviewers are assigned for only one week to the survey, leave to work on another project and then return a few weeks later.

Local high school graduates are often good interviewers. Since they tend to move out of the rural areas, it may be difficult to recruit these people.

University graduates may be technically qualified but frequently know very little about farming, dislike rural living, and have difficulty relating to the farmers. Consequently, the researcher must carefully evaluate these candidates attitudes and background.

Local teachers have often taught the farmers' children. They make excellent enumerators because they are educated, know the farmers, and are well known in the community.

Extension agents and other local officials may be capable of doing a good job, depending upon their reputation in the village. Also, respondents may not be willing to answer honestly questions about income due to a fear that such information will be passed onto the tax officers. Finally, answers to questions related to the enumerators official responsibilities may be biased, due to the respondents desire to be courteous and possibly to avoid reprisal.

C. Selecting Among Candidates

Once a candidate is identified, the individual should be interviewed to evaluate his or her suitability. Some researchers have successfully used a simple examination to test the candidates skill in the type of activities involved in survey work, including simple mathematical calculations, following instructions, writing ability, knowledge of local farming practice, and fluency in the language of the farmers to be interviewed. Based on this information, a team of interviewers are hired as trainees. During the training program, each participant is observed and the best ones are selected as interviewers (Flynn, 1976).

It is generally best to employ a small number of enumerators and extend the duration of the survey. The larger the number of interviewers, the greater the probability of bias being introduced as a consequence of differences in their personality, interest, and experience.

While it is useful to point out the desirable characteristics for interviewers, researchers often have little choice as to whom they hire. The "ideal" enumerators may not be available or an existing staff members must be employed. Consequently, careful attention must be paid to training the available individuals to conduct the survey in the desired manner.

D. Reducing Staff Turnover

It is important that all of the enumerators chosen to participate in the survey stay with the project until field work is completed. If interviewers leave the project prematurely, a new training program for the replacement staff will have to be conducted. This would be extremely time consuming and disrupt the field research.

In order to encourage the interviewers to stay with the project, a salary bonus may be offered. In addition to their monthly salaries all enumerators who continue with the project until it is completed are given a bonus equal to a specific percent (e.g. 25%) of the total salary earned. For example, if the enumerators are paid US$ 100 per month and the project lasts 6 month, the completion bonus would be computed as US$ 100 x 6 x .25 = US$ 150. This bonus would be forfeited if the enumerator quits before his contract is fulfilled.

IX. Training Interviewers

The purpose of a training program is to provide all enumerators with a common understanding of all aspects of the research. If this is accomplished, we can be confident that enumerator bias will be minimized. In other words, if we sent an enumerator to interview a farmer, he or she would obtain exactly the same response to every question as any other interviewer.

A. Length of the Training Program

The amount of training required will depend on several factors, including the length of the survey questionnaire, the experience of the enumerators, the amount of field supervision to be given, and the number of times each farmer is to be interviewed.

1) Length of questionnaire -- For a short questionnaire (1 hour or less) solicitating only closed answers and quantitative farm management data, a three to four day orientation would be adequate. On the other hand, a long questionnaire developed to obtain quantitative data plus information on attitudes through the administration of scales, and containing many open questions might require one to two weeks. Of course, the larger the size of the survey budget, the greater the absolute amount of time that can be justified for training.

2) Experience of enumerators -- The typical survey team includes some individuals who have a considerable amount of experience and others who are participating in their first survey. The training effort must be directed towards those with the least experience, and avoid the natural tendency to assume that since a few talkative trainees understand the points covered, the other also are following the discussion. Constant effort must be made to force each of trainees to participate and answer questions which will indicate their level of comprehension.

Particular attention must be paid to sensitizing inexperienced enumerators to recognize unacceptable answers to open questions. For example, when asked why he does not apply fertilizer, farmers sometimes respond with vague answers like "I don't want to use it". Enumerators must be trained to probe for the "real" answer to such questions such as "I don't have money to purchase the input". "I use traditional varieties that are not responsive to fertilizer" or similar concrete reasons.

Finally, the trainer should not assume that experienced interviewers know how to properly conduct an interview. Care must also be taken to observe if these trainees have acquired any poor interviewing techniques that are in conflict with the principles you are attempting to teach.

3) Number of rounds planned — If the questionnaire is to be conducted in several rounds (i.e. the interview schedule is divided and administered during more than one visit to the respondent), it is advisable to initially conduct a short training program, followed by review sessions at the beginning of each round. This strategy reduces the length of time and amount of information the enumerator must remember.

B. Focus of the Training Program

The training program must include both a formal orientation that can be conducted in the office and field training during which the enumerator attempts to apply the principals learned by actually interviewing several farmers.

1) Formal training — The specific topics which should be covered in a formal training program have been carefully outlined by Flynn (1976). These guidelines are extremely comprehensive and are quoted with slight modification.

- Details of the objectives of the research project and the role of the survey in accomplishing these objectives.
- Purpose and scope of the survey — to know the importance of their work, the functions to be performed and why and how the survey will be conducted.
- Understanding the system being studied — necessary in order to know what types of data that will be collected; the important relationships among data items and what constitute complete data.
- Definition of survey terms — to avoid confusion of terms and units of measure and to define the level of specificity required.
- Understanding the purpose and organization of all questionnaires to be used during the survey — to explain the importance of each instrument, to learn what specific facts are to be collected, where they go and in what form or units.

- Job and responsibilities of enumerators — whom to interview, how and when to interview, how to make entries, checking for errors and omissions, recalls and how to cope with uncooperative respondents.
- Special skills and techniques -- field measurement, yield estimation, and knowledge of units of measurement.
- Phasing of the survey — frequency and duration of weekly interviews, when situational or single visit forms will be administered, the agricultural calendar and implications for the work load.
- Administrative considerations — conditions of service, what is expected of the enumerator, payment arrangements, the chain of supervision, turning over questionnaires, supply of field materials, etc.

2) Field experience — After the team of interviewers has been oriented in the office, they should go into the field and practice interviewing. Initially, two or three group interviews of a single farmer should be conducted with the supervisor asking the questions and the trainees recording the answers. In a group discussion, these schedules completed by the enumerators can be used to illustrate the correct and incorrect way to record the farmers' responses. Next, each trainee is assigned several farmers to interview. Before implementing the survey, the project leader should observe each trainee conducting at least one interview and carefully check each completed schedule to see that it has been filled out correctly.

C. Level of Supervision

The amount of supervision to be provided will also influence the amount of training required. Suppose each of the enumerators will work alone in villages that are far from each other and difficult to reach because of bad roads. In such circumstances a relatively long training program will be necessary to guarantee that each interviewers have a complete understanding of the questionnaire.

On the other hand, if the enumerators are clustered in an area with good roads and the supervisor intends to make frequent support visits to check each interviewer's completed questionnaires, a shorter training program will be adequate.

In locations where frequent support visits are impractical, it is advisable to place at least two enumerators at each site. Then, if any questions arise they can consult each other.

X. Implementing the Field Survey

Once the enumerators have successfully completed the training program, the data collection phase of the research project can be initiated.

A. Obtaining Cooperation from the Village Leaders

The village is a very complicated social environment. In order to conduct social research, it is necessary to understand village dynamics. In general, strangers are treated with a great deal of suspicion and curiosity. Consequently, before beginning a survey, it is absolutely necessary to explain the project and get approval from local officials. This may include the police, the political leaders, and the extension office, depending on the country. As mentioned previously, the village leader interview is a useful way to explain the nature of the study to the local officials. During this meeting the purpose of the research should be explained. In addition, the researcher should emphasize to the leaders that the answers given by the farmers will only be used in this research project and will not be available to any government agency such as the tax office. In some cases, it might be advisable to actually interview the headman, although this schedule should be excluded from the data set during analysis — unless the leader was actually selected when the random sample was chosen. In addition, the local leaders should be kept informed as to the proposed research schedule and visited every time an interviewer returns to the village.

B. Obtaining Cooperation from the Village Farmers

While villages are generally willing to be interviewed, the researcher must make every effort possible to guarantee that the farmer actually reports the "true" answer to the question being asked. Several factors will influence the likelihood that the farmers will answer the questions to the best of their ability; including the researcher's knowledge of the village and farming situation before beginning the interview, the farmer's beliefs about the purpose of the research, the time the interview is conducted, the enumerator's behavior, and incentive payments.

1) Village orientation -- Before actually beginning to interview the farmers, it is advisable to have the researchers actually live in the village for at least one week. While an indepth understanding of village dynamics can only be gained through long term participant observer studies, this short village experience can still be valuable. First, researchers are typically educated and urban oriented. Being forced to live under the often meagre conditions of the village increases their sensitivity to and understanding of the problems faced by the villagers. Second, the time spent in the village before beginning the formal interviews will enable the enumerators to absorb information about the village that will enable them to administer the questionnaire more skillfully than would otherwise be possible.

In addition to the village orientation visit, the enumerators should live in the village for the duration of the survey. First, staying in the village will increase the interviewer's rapport with the farmers. Second, every experienced researcher has been faced with the problem of making sense out of the data, once it has been summarized back at the office. Since cost and time limitations make long term village residency unfeasible, living in the village while the survey is in progress will enable the researcher to obtain the indepth understanding of the farming system so necessary for meaningful analysis. This strategy represents a compromise between long term intensive studies and the frequent practice of arriving by car in the morning, conducting several interviews during the day, returning to the city at night, and repeating this procedure in another village the following day.[21]

2) Explaining the purpose of the research to the respondents — Before each interview begins, the enumerator must explain the purpose of the research to the respondent. It is important to emphasize that the farmer was selected at random. In addition the respondent should be told that the answers he gives are confidential and that only the researcher will see the completed questionnaire.

3) Timing of the interview — Since the farmer is generous enough to spend his time answering the researchers' questions, the interview should be conducted at the farmer's convenience. It is best to schedule the survey when there is little agricultural work in progress. In a rice area, convenient times might include after transplanting or after harvest.

---

[21] An interesting description of the value of long term village exposure is summarized in Srinivas, M.N., 1974.

If it is absolutely necessary to interview during peak work periods, the enumerator should ask the farmer what day and time of day is most convenient. The enumerator should never force the respondent to sit through the interview if he would rather be in his field. If forced to leave his work, the respondent is likely to rush through the questions and give inaccurate answers. Once an interview time has been scheduled, the enumerator should make all possible effort to arrive at the agreed upon time.

4) Interviewer's behaviour — The quality of the answers given is also influenced by the respondent's feelings towards the interviewer. Such factor as the enumerator's age, sex, dress, tone of voice, politeness, respectfulness, social class, and ethnic group can influence the farmer's willingness to answer the questions truthfully. Consequently, the enumerators must be sensitized to the importance of maintaining a good relationship with the farmer throughout the interview.

5) Rewarding the respondent -- Most researchers agree that the respondents should not be paid as this may bias their responses. Also, farmers not interviewed might be envious of the recipient causing tension in the village.

At the same time, there are several ways that the enumerator can express his appreciation to the farmers. Small gifts like a calendar may be given to the respondent before the interview. When the researcher does not plan to return to the village to evaluate change over time, such gifts as a small packet (1 kg) of an improved crop variety, or other inputs can be given to the respondent. Finally, many researchers have found that during the interview it is a good practice to offer the farmer a pack of cigarettes and keep a supply of candy to offer his children.

C. Supervising the Enumerators

After the enumerators are sent to the field, it will still be necessary to supervise their work, especially during the first few weeks of the survey. Unless errors are identified early, they will be repeated throughout the survey. If this happens, either the data collected will be useless or much time will have to be spent re-interviewing each respondent.

Supervision must be done by a senior researcher — preferably the project leader — who thoroughly understands the research goals and each item in the questionnaire. Attention must be paid to both methodological and logistic details, as outlined by (Flynn, 1976).

First, supervision is necessary to:

- continue the training process, reemphasizing difficult aspects of the questionnaire.
- assist in handling interviewing problems such as refusals.
- check questionnaire for errors, omissions, and illegible answers.
- provide moral support.
- ensure that the enumerator is actually interviewing the farmers.
- ensure that the enumerator's relationship with the farmers is good.
- collect completed interviews.

If any problems are observed, the supervision must immediately explain the error and instruct the enumerator how to avoid repeating the some mistake. If errors are found, the interviewer must immediately return to the farmers and ask the question again in order to obtain the correct answer.[22]

In addition to on-the-spot checking of the questionnaire, it is advisable to select a subset of critical questions (for example, hectarage and production) and have the field supervisor reinterview a 10% sample of the respondents regarding these items. If the same answer is given a second time, the correlation coefficient between the first and second response will be 1.00. In this way, it is possible to measure the reliability of the data on these essential variables and gain a feeling for the reliability of the data being collected.

Careful review of the questionnaires is not only necessary to identify interviewers' errors, but also encourages the enumerators to be more careful in recording the data. Since they know that someone is really concerned about accuracy, the enumerator will try to meet the supervisors' expectations.

---

[22] For many types of analysis, including regression and discriminant analysis, a missing value for one variable makes it necessary to drop the case from the sample, reducing the total number of cases. For example, if you want to estimate a six variable regression model and 25 of 100 sample farmers failed to answer any one of the six critical questions, the sample size would be reduced to 75 respondents.

Second, the supervisor must check on various aspects of logistic support to see that the enumerators have:

- an adequate supply of questionnaires
- an adequate supply of pens, paper and other special items.
- special clothing if required such as boots, umbrellas, and a waterproof container to protect the questionnaire from the rain.
- good means of transportation — usually a bicycle which is kept in good repair, and
- are paid at regular and predetermined intervals.

Field supervision, can also be facilitated if all interviewers are required to keep a daily diary. Every day, the enumerator should enter notes about any problems encountered, difficulties in asking certain questions, interesting observations about the local farming system, and similar items. When the supervisor visits the village, the diary should be reviewed and all points discussed.

Visiting the interviewers frequently will not only improve the quality of data collected, but also increase the supervisor's knowledge about the local farming situation. Consequently, the indepth understanding gained through these visits will enable the supervisor to better interpret the data collected.

D. Special Problems in Field Implementation

A number of problems are likely to occur during the survey which merit detailed discussion.

1) Rewording of questions -- Once the enumerators become familiar with the survey schedule, they often try to paraphrase the question — using their own words. Interviewers should be instructed to read each of the questions exactly as written, If the questionnaire is properly pre-tested, it will seldom be necessary to modify the item. Unless the same wording is used to present each question to every respondent, it will be impossible to compare farmers'responses. While a supplemental explanation may be necessary in some cases, these should be kept to a minimum.

2) Question order — For various reasons, the interviews sometimes decide to ask the questions in some order other than as presented in the questionnaire. As explained in the discussion on questionnaire development, deviating from the prescribed order may bias the farmers' answers. Consequently, the enumerator must ask all questions in the sequence presented in the questionnaire.

3) Speedy interviews — As the interviewers become familiar with the questionnaire, there is a tendency for them to rush through the questions. Since accuracy — not speed — is the primary objective, the supervisor should check to make sure that all interviewers take approximately the same amount of time for each schedule. This can be easily monitored by instructing the interviewer to note the starting and finishing time on the first page of the questionnaire. Interviewers should never be paid according to the number of schedules completed, since such as system tends to reward speed over accuracy.

4) Inappropriate respondent — Once the sample has been selected, the survey team must make a faithful effort to interview the chosen respondents. After beginning the interview, it may become obvious that the individual should not have been selected in the first place. For example, the sample should included only irrigated farmers and the individual's farm is rainfed. In such instances, the interview should be terminated and a replacement selected.

5) Refusals — After being contacted a respondent may refuse to be interviewed. All effort should be made to explain the nature of the project and that he has been specially selected by a random process. Also, the interviewer should offer to return at a time which is convenient for the respondent. A refusal may indicate that the interviewer was disrespectful to the farmer. Hence, before dropping the farmer from the sample, the supervisor should attempt to obtain his cooperation. If this approach fails, another respondent should be chosen. If the farmer is forced to cooperate by the village leader, the quality of the interview may be very poor. In actual practice, refusals are extremely uncommon.

6) Missing respondent — The most critical non-response problem arises when the interviewer can not locate the selected respondent. For example, he may be out of town visiting a relative or working in the city during the slack season. It is a serious mistake to simply choose another farmer. It is likely that individuals who migrate out of the area in search of employment are quite different — in terms of non-farm employment, for example, than continuous residents. Exclusion of such individuals may greatly distort the research results. Also, interviewers should never complete the interview by asking the questions to the absent respondent's wife, son or any other available person. Such action could seriously reduce the accuracy of the information. The best strategy to handle this situation is to find out when the respondent will return and try to reschedule the interview after that time.

A rule should be established to guarantee that all "missing respondent" problems are handled in the same manner. For example, you may decide that the respondent is replaced only after three attempts are made to interview him or if he will not return within a certain time period (i.e., six months). Dropping a few such cases will not seriously effect the research results. On the other hand, if more than 10-15 percent of the randomly selected respondents are unavailable, the survey timetable should be revised to provide an opportunity for the enumerator to return to the village at a future date to interview these individuals.

XI. Choosing the Data Analysis System

The type of analysis to be performed should have been decided when the project was being planned, since alternative analysis strategies have different data processing implications. When the type of analysis to be performed is explicitly considered at an early stage, it is possible to introduce procedures which will reduce the time required to process the data. For example, if computer analysis is planned, the researcher would probably want to precode and prenumber the questionnaire.

Alternative strategies for analyzing survey data may be compared in terms of time required to prepare the data for analysis, time required to execute the desired statistical operations, experience needed to utilize the system and type of analysis which can be executed, as summarized in Table 2.

Table 2. Characteristics of selected data analysis systems

System	Data file preparation	Statistical operations	Experience required	Analytical possibilities
Tabular sheets	rapid	slow	minimal	minimal
Sorting strips	rapid	slow	minimal	minimal
Computer				
Fortran	moderate	moderate	very specialized	unlimited
SPSS	slow	rapid	specialized	unlimited

A) Tabular Sheets

The simplest method for preparing data for analysis is to use a large sheet of lined paper. First, the researcher should decide what is the most meaningful way to order the respondents, based on the type of inter-group comparisons to be made. For example, if you intend to compare villages, the questionnaires for all respondents residing in each village should be grouped together so that when the farmers' responses are transferred onto the data table, the respondents from each village will be listed in sequence. Alternatively, the researcher may choose to group the respondents in terms of irrigation class, farm size, yield, tenure, variety grown or any other relevant characteristic. In some instances the researcher may wish to stratify the respondents within each primary group. For example, all of the respondents could first be ordered by village. Then, within each village respondents might be grouped by irrigation type. Finally, within each irrigation class, the farmers using modern varieties may be listed first, followed by those planting traditional varieties. The decision as to how to array the respondents must be based on how the researcher plans to present the data in the basic analytical tables. Careful planning as to the order in which the respondents are to be arrayed in the data table will greatly reduce the time required to compute the required statistics and prepare the desired basic tables. Second, each respondent's name or code number is listed down the most left hand column and the name of each variable is entered across the top row of the sheet. Third, the value for each variable for each farmer is transferred from the questionnaire into the appropriate column and row. Once all of the data has been entered, the researcher can proceed with analysis.

When tabular sheets are used, the data can be rapidly prepared for analysis since it is only necessary to transfer data from the questionnaires to the tables.

When tabular sheets are employed, statistical operations must be done manually, usually with the assistance of a simple electronic calculator or a programmable model. Consequently, analysis is time consuming since each piece of data must be punched

into the calculator every time a statistic is to be computed. As the sample size and number of variables increase, the time required to analyze the data also increases. Hence, this method is most appropriate when the sample size is small (less than 100 cases) and the number of variables to be evaluated can be listed across one or two tabular sheets.

On the other hand, researchers with little or no previous experience can easily utilize this system since it is simple to learn. Also, since all of the values for each variable are listed in a column, it is easy for inexperienced researchers to spot and correct extreme variable values which were erroneously entered into the table.

The greatest disadvantage of preparing the data in tables is that only limited analysis such as estimating sample means and frequency counts may be easily performed. Since there is no way to easily separate subgroups within the sample, it is very time consuming to even compare subsample means such as irrigated yields vs. rainfed yields or fertilizer users vs. nonusers — other than those which were used to order the respondents in the table. If regression is to be used, it will be necessary to punch into a programmable calculator the values of the dependent variable and all of the independent variables each time a different model is to be estimated.

In summary, tabular data preparation is most appropriate when the analytical objective of the research project is to calculate descriptive statistics for the total sample or a few specific subgroups; and the sample size is relatively small. In such a circumstance, it is possible to complete data analysis within a short period since little time is required for data preparation.

B) Sorting Strips[23]

Sorting strips represent one step, in terms of sophistication, above tabular sheets. In fact, each strip is simply the equivalent of one column on the tabular sheet which has been separated from the table. Sorting strips may be easily made from heavy weight paper cut

---

[23] For a detailed description of this system, see Taylor, Donald C., 1973.

into 2" x 20" strips[24]. To divide the strip into cells, first a line is printed lengthwise through the middle of both the front and back side of the strip. Next, vertical lines are printed at one inch intervals on both sides of the strip. Then, each of the resulting cells is numbered in sequence, 1-40 on the front and 41-80 on the back. It is useful to have strips printed on paper of several different colors. By assigning all respondents with a common characteristic to be used as a basis for analysis (i.e., all farmers living in each village, those with a common irrigation class, etc.) to a single color strip, data manipulation will be somewhat easier. Before any data is transferred to the strips, a codebook must be prepared. This is simply a reference guide in which each question is numbered in sequence followed by a short description of the item, the strip number (when more than one strip is required) and the number of the cell into which the data is to be transferred. The questionnaire of each respondent must be assigned a unique identification number so that the respondent can be easily identified if it is necessary to refer back to the original questionnaire. When preparing the codebook, the respondent's ID number should be assigned to the first cell of the sorting strip. If more than one strip is required to accommodate all of the variables, cell 2 of each strip must be reserved for the strip number in order to permit the researcher to be able to determine which strip is associated with which variables.

The 38 remaining cells on the front side of the strip are assigned to variables 3 through 40 (variable 1 and 2 are the farmer's ID and strip number). Variables 41-80 are assigned to the back side of the first strip. Additional strips may be used until all variables have been allocated a position, but at the beginning of each strip the farmer's ID number and strip number must be recorded. If, for example, each questionnaire contains 160 variables (i.e., 160 questions), then two strips will be required per questionnaire. Once the codebook is completed and each variable has been assigned to a sorting strip and cell, the value for each variable in the questionnaire is simply transferred to its preassigned position.

---

[24] Sorting strips may be made in any length and width desired by the researcher. The suggested dimensions produce a strip that is particularly easy to handle.

The researcher may not want to transfer each variable in the questionnaire onto the sorting strip. In some cases, it may be best to first summarize the data and then transfer aggregated variables such as gross returns, net returns, total input costs, or mandays per farming operation onto the sorting strip. After all desired variables have been entered onto a set of strips for each farmer, the researcher can proceed with analysis.

If sorting strips are used, the time required for data preparation will be slightly greater than when tabular sheets are prepared. Yet, if the researcher carefully selects and transfers only the variables to be actually used in the planned analysis, the time required for sorting strip preparation can be significantly reduced. Hence, the researcher should be ready to analyze his data soon after the survey is completed.

In terms of time required to execute statistical operations, there is no difference between tabular sheets and sorting strips. On the other hand, when the data is on strips, it is very easy to manipulate the data in order to compare the characteristics of two or more subgroups with each other. For example, suppose the researcher wants to determine if there is a statistically significant yield difference between rainfed and irrigated farms. Assuming the irrigation class variable is coded in cell six (rainfed = 1, irrigated = 2), and yield is entered in cell 12, the researcher simply sorts all strips with 1 in cell 6 into one pile and all strips with 2 in cell 6 into another pile. Next, the mean and standard deviation are calculated for each subset of respondents and the difference between means is evaluated using a t-test. The actual statistical analysis may be made using either a hand held calculator or a programmable calculator. Yet, as the sample size increases it becomes quite cumbersome to manipulate the strips. Consequently, this system is most appropriate when there are no more than 100 cases in the sample and a few hundred or less variables are coded.

Sorting strips have the additional advantage of being easy to learn to use. After a short training session, someone with no previous experience will be able to skillfully manipulate the data and rapidly make inter-group comparisons.

While sorting strips enable the researcher to easily compare characteristics of subgroups within the sample, the researcher is still limited — in terms of analytical alternatives -- to calculating such basic statistics as means, standard deviations and correlation coefficients. As with the tabular system, in order to estimate regression models, the researcher must manually punch into a programmable calculator the values of the dependent and all independent variables for each model.

In general, the main advantage of sorting strips, compared to the tabular system, is that strips facilitate the manipulation of the cases in the sample. When the sample size is relatively small and there is a need to generate descriptive statistics for subgroups of the sample shortly after the survey is completed, the sorting strip approach has considerable merit. On the other hand, if the sample size is large and the researcher wishes to utilize advanced statistical procedures, this system is cumbersome.

C) Computer Analysis

In contrast to the tabular and sorting strip systems, if a computer is to be used, the data must be prepared according to a precise set of rules. Data preparation involves entering the values for each variable onto computer cards or a magnetic tape. If a precoded and prenumbered questionnaire was used, this can be done directly from the questionnaire. On the other hand, if the questionnaire was not prepared for direct punching, the researcher will have to transfer all of the data onto coding sheets. Before this can be done, it will be necessary to develop a codebook, indicating the coding sheet and columns in each coding sheet into which each variable value should be entered. Using the codebook as a guide, all of the data from the questionnaire must be transferred onto coding sheets and then punched onto cards (or a magnetic tape). Preparing data for computer analysis is generally quite time consuming, especially if a precoded and prenumbered questionnaire was not used.

If a language like Fortran IV is to be employed, the coded data can be analyzed by means of a program which instruct the computer to execute desired operations. On the other hand, if a system of prepared computer programs is to be used, it will be necessary to first set up a systems file

according to the specifications of the package. Generally, file preparation is extremely time consuming since all of the required words and characters must be used in accordance with a many specific rules in order to successfully establish the systems.file.

One advantage of computerized analysis with systems of prepared programs is that once the data and data file have been prepared, statistical operations can be executed rapidly.

One of the disadvantages of using the computer is that the researcher must first learn how to instruct the machine. While computer languages are much more difficult to learn to use than prepared programs, even the prepared systems of programs like SPSS require a considerable investment of time. At least 2-3 weeks of study will be required before the inexperienced researcher will have acquired sufficient knowledge to begin to analyze the data.

Most relationships in social science are influenced by several compounding factors. The greatest advantage of using a computer is that it opens up a wide range of analytical possibilities which would be nearly impossible to take advantage of if data tables or sorting strips are used, including multiple regression, and more sophisticated methods of analysis. In addition, the computer may also be programmed to rapidly compute all descriptive statistics and perform arithmetic operations (add, subtract, multiply and divide) on the raw data in order to create new variables. Since both advanced and descriptive statistics are calculated by a machine, the researcher can be confident that no computational errors have been made. Finally, a computer produces a printed record of the analysis in an easily readable format that may be saved for future reference.

In summary, the computer is a valuable research tool because it enables the researcher to rapidly execute a wide range of statistical procedures. Consequently, the researcher is able to rigorously test hypotheses which could not be otherwise investigated. In addition, when the sample size and number of variables to be evaluated are extremely large, the researcher may choose to use a computer — even if only

descriptive analysis is planned. On the other hand, data and file preparation is time consuming, so the results may not be available for several months. Also, if the researcher has no computer experience, it will be necessary for someone to invest a considerable amount of time to learn a computer language or rules governing the use of the prepared system of programs. If a computer is not available locally, much time will be spent travelling to and from the terminal since each analytical job usually must be run several times before programming errors are completely eliminated. Finally, computer time is often extremely expensive. Consequently, the researcher must consider the cost of analyzing the data with a computer — compared to using a manual systems — when weighing the advantages and disadvantages of using this analytical tool.

XII. A Brief Introduction to Computer Analysis

After the decision to employ a computer has been made, the researcher must still decide which type of machine best fits the research projects' needs. In addition, a specific analytical language or package must be selected. This section briefly introduces some of the available alternatives. Since most social scientists will find the SPSS system of prepared programs most suitable for their needs, a somewhat detailed description of this system is included.

A. Hardware

There are two types of computers that may be used to analyze survey data.

1) Mini-computers — Several types of mini-computers are available. These machine are superior to the desk top programable calculator because once the data is entered into the memory, it can be stored on a tape and recalled for future analysis. This permits the researcher to both manipulate the data (i.e., perform transformations, select data subsets, etc.) and do statistical analysis. Many of the available models will accommodate the same type of analysis as the standard computer, but have a smaller data storage capacity. In addition, usually a special language must be learned in order to program these machines.

2) Digital computers — The analytical capability of a large computer like the IBM-360 is practically unlimited. In situations where such a computer is available and in-depth analysis is planned, the researcher should consider using the digital computer.

B. Softwear

Computer softwear may be grouped into language like FORTRAN, BASIC, and COBOL, and prepared systems of programs packages such as SPSS, BIOMED, and SAS.

1) Languages — In order to analyze data through a computer language, each step in the statistical procedure must be completely delineated using alphabetic characters, numbers, key words, and certain special symbols. For example, in using FORTRAN to determine yield-given production and area planted — you would first outline the problem as illustrated in the flow chart of Figure 1.

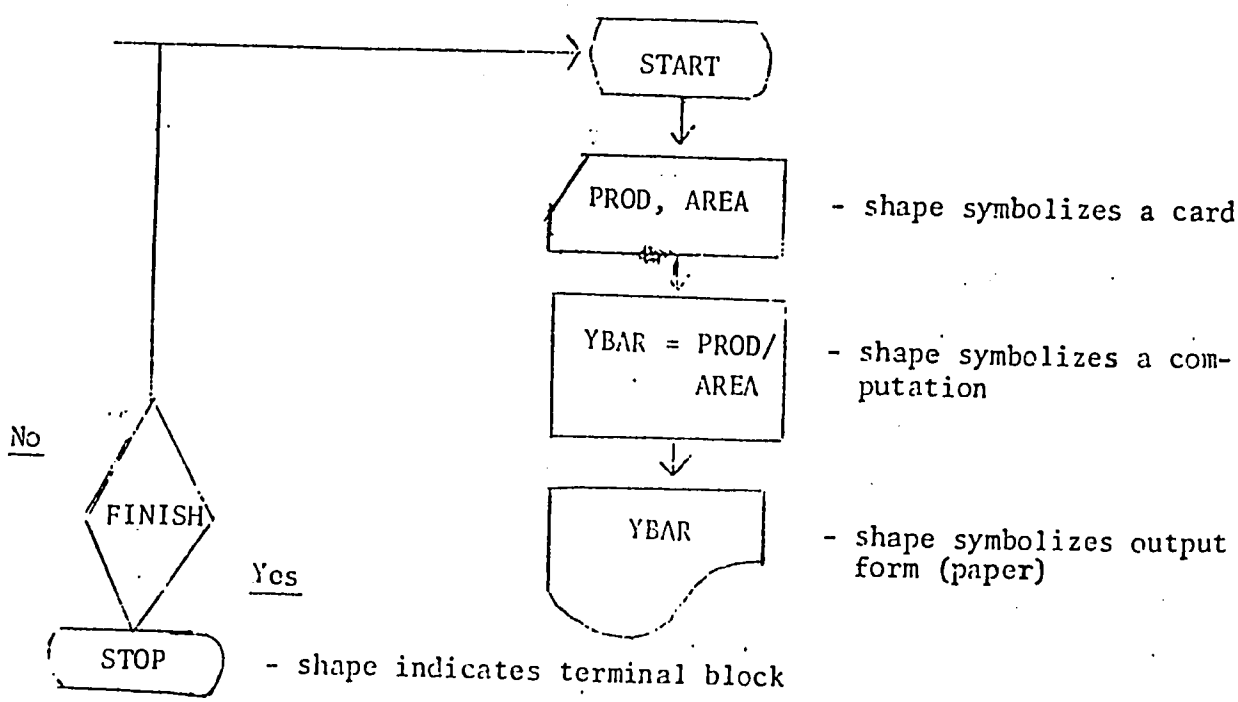

Figure 1. FORTRAN flow chart for calculating yield.

Once the analysis problem has been outlined, instructions must be written in a form which the computer can interpret. For example, let us assume a programmer has a set of data with production figures coded on the second card, columns 29-31, and area farmed in colums 32-34. To calculate yield, the following program would be required:

FORTRAN Statement	Meaning
READ (5,100) PROD, AREA	The computer must read data from media 5 (cards). 100 is the code for the format statement.
YIELD = PROD/AREA	Yield is computed by dividing production by area planted.
WRITE (6,105) YBAR	
105 FORMAT (F7.2)	The output is in real number (F). There are 7 spaces involved, a sign and decimal each occupying one space and five numbers.
100 FORMAT (/28X, F3.2, F5.2)	The data appears on the second card. The first 28 spaces are skipped. Production data is in columns 29-31 and area farmed in columns 32-36.

The main advantages of a computer language are that selected variables can be readily chosen for analysis. Also, a program can be written to execute virtually any desired operation. Finally variables can be accessed immediately since it is not necessary to first prepare systems files as is required when using prepared program systems.

On the other hand, since the instruction statements are relatively complicated, a researcher must invest a great deal of time learning the language or depend on someone else for programming. Yet, it is impractical to rely on an assistant for programming when a large number of models are to be estimated. Not only will a great deal of time will be spent writing instructions, but also if someone unfamiliar with the data does the programming, errors are more likely to occur.

2) Prepared system of programs — These systems have been developed in order to reduce the amount of time required to program the data manipulations and statistical operations most frequently used by social scientists. Because packages are preprogrammed, once the data and systems files are created, words and symbols can be used to "talk" to the computer. For example, after a SPSS systems file has been set up, the mean yield of a sample could be estimated with the simple statement:

```
COMPUTE                    MEANYIELD = PROD/AREA
1                          16
```

Compared to the FORTRAN statement, this is quite simple. Since prepared systems of programs like SPSS use key words which enable the researcher to execute statistical operations without detailing every step in the procedure, they are relatively easy to learn and use.

C. The Statistical Package for the Social Sciences (SPSS)

A detailed description of SPSS may be found in the programming manual written for this system of programs[25]. The discussion that follows is intended to illustrate how easily anyone with a high school education can learn to use a prepared program. By stripping away some of the mystery and awe that surrounds the computer, it is hoped that more researchers will be encouraged to increase the rigor of their analysis through the judicious use of this research tool.

1) The codebook — If an uncoded questionnaire was used, it will be necessary to first prepare a codebook to guide the data processors in coding the answers to each question. Coding involves transferring data from the questionnaire to coding sheets. The codebook also serves as a reference when analysis is performed. For each variable, the question number is noted and an abbreviated text of the original question is written. This is followed by a variable name associated with the variable[26] and the number of the card and columns on the IBM code sheet into which the response should be entered[27]. In addition, the values assigned to each qualitative response category are noted as illustrated below:

---

[25] Nie, et.al., 1975.

[26] When setting up the data file, an acronym of no more than eight letters is assigned to each variable. During analysis, this word is used to access the data associated with that variable.

[27] Each card has 80 spaces into which the value of variables may be entered.

Question No.	Question	SPPS variable Name	Card (No.)	Columns (No.)
1	Farmer Identification Number	IDNO	1	1 2 3
2	Card Number	CARD	1	4 5
3	Farmer's total rice area	AREA	1	6 7 8 9
4	Farmer's total rice output	PROD	1	10 11 12 13
5	Farmer's tenure	TENURE	1	14 15

Code

1 = owned 3 = leasehold
2 = share tenant 4 = other

If a precoded and prenumbered questionnaire was used, the data may be directly punched onto cards (or magnetic tape) — eliminating the need to transfer the data onto coding sheets. Since the original questionnaire was precoded and prenumbered, it may be used as the codebook by simply adding the name of the variable associated with each variable.

2) Coding Conventions — When developing the codebook for a questionnaire that was not precoded or when designing the precoded and prenumbered questionnaire, several points should be kept in mind. The following guidelines are not a complete set of instructions for coding data for computer analysis when cards are employed. Rather they are practical lessons learned through experience which are often overlooked by inexperienced researchers.

It is essential to identify each physical card in a data set uniquely. This can most easily be done by punching the card number

and observation number (respondents ID number) on each card. In most data sets they would require no more than 6 columns. The same 6 columns should be reserved as ID columns in every card — usually the first six columns on the card.

A coding sheet has 80 columns and 30 rows which correspond to 30 computer cards[28]. In other words, each card is represented by one row on the coding sheet. It is best to code all variables which occupy the first 80 columns on the first coding sheet for farmers 1 through 30. Then, data for farmers 31 through 60, etc. are entered on the succeeding coding sheets. All of the coding sheets which include the first 80 columns of data may be called a set of coding sheets. Since more than 80 columns will be required to enter the values of the variables to be analyzed, additional sets of coding sheets are used. The second (third etc.) set of coding sheets correspond to the next 80 columns of data. It is a very poor practice to code all of the data for each farmer on a single sheet. If this is done, it will be impossible to look down the set of columns designated for each variable in order to identify numbers placed in the wrong column or extreme values.

Some researchers believe that all 80 columns in <u>each coding sheet</u> should be completely filled with a number or zero because key punch errors are more likely to occur if internal blanks are left in the coding sheet. For example, if five spaces are reserved for yield, but the yield was only 5200, it should be coded 05200. The disadvantage of this procedure is that if blanks are not left between variable values, it is difficult to check for misplaced or extreme variable values since the data is run together. Yet, by simply drawing a vertical line down the coding sheet between variable values, it will be possible to separate the variable values for checking.

Each computer program has a specific way of handling <u>blanks</u>. SPSS treats a blank as zero. If the column into which "yield" is to be entered is left blank because the data were not collected for one

---

[28] Letters, numbers or special symbols may be entered in each column. Each of the 80 spaces correspond to one column on the standard computer card.

observation (respondent) the computer will read it as zero – biasing any statistical estimate . SPSS allows the researcher to designate three different types of missing data. For example, it may be desirable to distinguish between missing data arising when the farmer refuses to answer, the farmer doesn't know the answer, or the enumerator fails to record the answer.

All variables should be entered in the codebook (and coding sheets) in the order in which they appear in the questionnaire. Rearranging variables into a seemingly more "logical" order will greatly increase both the amount of time required to code the data and also the probability of making coding errors.

If two variables have the same response categories (e.g., tenure status of rice land" and "tenure status of vegetable land"), confusion will be reduced if the same code number is associated with each response category. For example, you should never use 1 to designate "owner" on rice land and 1 to designate "tenant" on vegetable land.

Several coding forms will be required to transfer all of the data from all the questionnaires, since each coding sheet has only enough space for 80 numbers. The quickest and most accurate procedure is to code the 80 columns on the first page (corresponding to the first computer card for each respondent) for all of the respondents in the sample, using 2 pages for a 60 farmer sample, 3 for a 90 farmer sample, etc. Then, fill in the 80 columns on the second page (corresponding to card 2) for all farmers. Continues in this manner until all the data is transferred to the coding forms, If several persons are available, it is best to assign each person a specific page (card) to code for all respondents.

Once the data has been entered on the coding sheet, each variable should be checked by following the set of columns assigned to each variable down the coding sheet. In this way, data coded in the wrong column and extreme values can be quickly identified and corrected.

It is advisable to randomly check the data entered by each coder. This can most easily be done by having a second person recode every 10th card (10 percent check) and compare these values to the original figures.

Each researcher must decide how much error can be tolerated. For certain critical variables such as yield and physical input levels, it may be advisable to check the values for each respondent.

3) The Input Medium — Once the data has been checked, it is ready to be keypunched and verified -- a process whereby the data are essentially repunched to check for errors. Usually, these operations are done by skilled personnel, paid on a per record (card) basis.

Data may be stored on either cards or tape. Since virtually all computers accept data on cards, it is the most widely used medium. On the other hand, when cards are utilized, it is necessary to feed all data cards into the computer each time analysis is performed. After repeated use, the cards tend to fray — a process which results in the rejection of the card by the computer. Consequently, these frayed cards will have to be repunched. In addition, a large set of data on cards is very bulky. Finally, cards are likely to become mixed up after continued use, and if stored where the humidity is high, they will deteriorate.

It is possible to keypunch data directly — from the questionnaire or coding sheets -- onto tape. Once this is done, it is no longer necessary to feed the data cards into the computer each time analysis is performed. The computer simply picks the data from the tape when the program is submitted. In addition, if cards are required at some future date, it is possible to use a utility program to output a set of cards from the taped data. On the other hand, if a set of cards are initially punched, it is also easy to subsequently transfer the data onto tape via a utility program.

For most situations, it is best to punch data onto cards and then use the cards to produce a data tape. This is advisable for several reasons. First, coding errors on cards can be easily corrected by the researcher on an ordinary keypunch machine. Errors on tape must be corrected on a key-to-tape machine by a skilled operator. Second, all computer systems accept cards, but many don't accept taped data. Third, it is easy to accidently erase a data tape. Hence, a set of cards serves as a back-up should the tape be erased.

If the researcher decides to tape the data and the SPSS systems file, a duplicate of the original tape should be made immediately, since almost every researcher will at sometime in his career accidently erase a data tape. This is particularly critical when data is punched directly from a prenumbered questionnaire to tape, since accidental eraser would require that all the data be repunched. Also, if any modifications are made on the initial SPSS systems files, the researcher should immediately duplicate these new files so that if they are accidently erased, it will not be necessary to go through the lengthy process of reconstructing them.

4) Checking the data — After the data have been keypunched on cards or tape, an utility program should be used to obtain a list printing of the data before it is used for analysis. This printout must be carefully checked against the original data as it appeared on the data coding sheets (or questionnaire if the data was directly punched). Also, the keypuncher may have skipped rows on the coding sheet or left complete coding sheets unpunched. Hence, it is important to count the cards that are returned by the keypuncher and compare this total to the number of rows of data submitted for keypunching. In addition, the ID information on every card should be carefully checked against the numbers on the original coding sheets in order to catch punching errors. Finally, the researcher may wish to check a randomly selected set of computer cards against the coding sheets in order to estimate the rate of keypunching error.

It is also possible to perform a range check of the values in the data with editing programs. After specifying the acceptable range of values for each variable, the program scans the data and notes all values outside the specified range. For example, suppose you feel that yield should not exceed 10,000 kg/ha. Once this maximum is specified, the utility program will flag values greater than 10,000. By using this program, extreme yields such as 18,000 kilograms can be caught and corrected before analysis is started.

5) Establishing a SPSS systems file — Before it is possible to manipulate the data using the SPSS system of prepared programs, a systems file must first be created. This can be done as soon as the researcher has identified the card and the columns on each card into which each variable will be coded.

Control cards are used to instruct the system how to process the data. Each SPSS control card has a "control field" (columns 1-15) for entering control words and a "specification field" (columns 16-80). Only control words like VARIABLE LIST, INPUT FORMAT, VARIABLE LABEL, VALUE LABEL, FREQUENCES, REGRESSION, and RECODE may be entered in the control field. These words or set of words inform the system as to the type of specifications that will follow so the system will know how to read, interpret and respond. In order to establish a SPSS system file, the data must first be defined. Several of the most important data definition cards are VARIABLE LIST, INPUT FORMAT, VALUE LABEL, and MISSING VALUES.

VARIABLE LIST — In order for the computer to recognize each piece of data, each variable must be given a unique name of no more than eight numeric or alphabetic characters. These names must be entered in the codebook for reference during programming. Next, these names must be listed on the VARIABLE LIST card in the exact order they appear in the coding sheet. For example, if we use the variable name IDNØ for farmer's identification number, AREA for hectares planted and ØUTPUT for total production, the first variable list card would appear as follows:[29]

```
VARIABLE LIST IDNØ, AREA, ØUTPUT
1             16
```

If more than one card is required to list all variable names, the first 16 columns on succeeding cards are left blank and additional variable names are entered in columns 16-80.

These names will be used during analysis to identify the variables upon which statistical operations are to be performed. Therefore:

- it is best to use only alphabetic names which can be logically associated with the underlying variable (i.e. AREA for area planted, TEN for tenure class etc.)
- use as few characters as possible to name each variable since these letters will have to be punched every time you use the variable.
- when preparing program cards, the letter 0 should always be written as Ø to distinguish it from the number zero.

---

[29] The number written under the letters indicates the column into which it must be punched.

INPUT FORMAT — In order for the system to locate the numbers associated with each variable name, INPUT FORMAT cards must be prepared. These cards identify the columns in which the values for each variable were entered. Assume that the data set includes only three variables IDNØ, AREA and PRØD which require three, four and four digits, respectively. To express this information, the INPUT FORMAT card would be written as:

```
INPUT_FØRMAT___FIXED_(F3.0,F4.1,F4.0)
1                16
```

The figure in front of the decimal indicates the number of columns occupied by the variable and the figure following the decimal indicates the number of decimal places in the variable.

VARIABLE LABELS — Often it is difficult to remember the meaning of each of the variable names created. To assist the researcher, each variable included on the VARIABLE LIST card may be given an extended name of up to 40 characters. Whenever the variable is used in analysis, the extended name will be written on the computer printout. Suppose we want to clarify the meaning of the variable label FERT, used to identify kilograms of fertilizer input. To associate an extended name with this variable, the VARIABLE LABELS card would be written as:

```
VAR_LABELS... FERT_TØTAL_KGS_UREA_AND_PHØSP
1              16
```

VALUE LABELS — When characteristics like tenure are coded as discrete values (i.e. 1 = owner; 2 = tenant) a VALUE LABELS card should be prepared. By including these cards, the program will print out the label associated with each value whenever the variable is used in analysis. For example, a VALUE LABELS card for the variable "TEN" (tenure) would be prepared as follows:

```
VALUE_LABELS__TEN(1)ØWN(2)LEASE
1              16
```

MISSING VALUES — No matter how careful the researcher checks the questionnaires in the field, the answers for some questions will be missing. The manner in which these missing values have been coded must be documented on the MISSING VALUES card. Up to three different types of missing values may be declared for each variable. For example, the researcher may want to distinguish between blanks (i.e. no answer recorded by the interviewer), absurd values that were deleted by the researcher, and cases where the farmer refuses to answer the question. The declaration of missing values for production data is illustrated below:

```
M I S S I N G _ V A L U E S _ Ø U T P U T ( - 9 , - 8 , - 7 ) /
1                                 16
```

Negative values may be used in order to avoid confusing the number used to designate missing values with the value of an actual answer. When negative numbers are employed, two columns must be set aside for each variable[30].

4) Data Modification — A number of procedures can be employed to transform the original data, once it has been entered into a SPSS systems file by using such cards as COMPUTE and IF.

The COMPUTE card permits the researcher to, for example, add, subtract, multiply and divide variables; and take the square root, log and square of a variable. The resulting new variables can then be used in the same way as the set of variables included in the original SPSS systems file.

The IF card allows the researcher to define new variables or to alter the values of existing variables by specifying logical statements and arithmetic expressions. Such relational operators as greater than (GT), less than (LT), and equal to (EQ) may be used here. Suppose we want

---

[30] If a precoded and prenumbered questionnaire is used, this convention must be incorporated into the questionnaire design. The meaning of each missing value should be noted in the codebook.

to compare small and large farms. We can use IF cards to create a new variable FM which includes two categories — farms with a total area of less than 1.0 hectares and farms greater than 1.0 hectares — as follows:

```
IF ... (AREA LT 1.0) FARM = 1
1     16

IF ... (AREA GT 1.0) FARM = 2
1     16
```

5) Data Selection — A particularly useful feature of SPSS is the ease with which cases meeting certain conditions may be selected for analysis. For example, it would be possible to select all rainfed farmers using more than 50 kgs of nitrogen and over 40 years of age with the SELECT IF card. Other data selection cards permit you to choose a random sub-sample (SAMPLE) and weight the cases (WEIGHT).

6) Procedure Cards — SPSS provides the user with a great variety of procedures, some of which are described in Table 3. In addition, each procedure offers a comprehensive set of statistics and options which may be requested by the researcher.

Table 3. Some frequently used procedures available in the SPSS system of prepared programs

Procedure Name	Description
CONDESCRIPTIVE	Estimates mean, range, maximum and minimum values.
FREQUENCIES	Constructs one-way frequency tables and histograms.
CROSSTAB	Estimates joint frequency distribution of up to three classificatory variables.
BREAKDOWN	Computes sum, mean, standard deviation and variance of variables among subgroups.
T-TEST	Compares sample means.
PEARSON CORR	Computes the correlation between two variables.
SCATTERGRAM	Makes a two dimensional plot of data points.
REGRESSION	Estimates the relationship between one dependent variable and more than one independent variable.
ANOVA	Provides a one-to-five way analysis of variance and covariance for factorial designs.
DESCRIMINANT	Estimates the relationship between a discrete dependent variable and several independent variables.

7) Miscellaneous topics related to using SPSS.

Job Control Language — In order to run a job on a computer, it is necessary to utilize Job Control Language (JCL) cards to instruct the computer as to the program you are using, core space required, and similar details. Selection of the correct JCL cards is unique to each computer and may be initially confusing. Consequently, the beginner will need to consult an experienced user to learn about JCL conventions for the local computer installation.

Editing the SPSS systems file. —Once all variables have been listed, named, given values and assigned missing values, the file is ready to be edited. This is accomplished through the use of the EDIT control card which instructs the computer to check all control cards to determine if they are internally consistent and correctly keypunched. The output from the EDIT run identifies each error and specifies the problem. Once all errors have been corrected, the file is ready to be taped or used in card form.

Modifying a SPSS systems file -- It is possible to modify a SPSS systems file once it is established. Details are outlined in the SPSS manual. If cards are used, the original cards are simply replaced by new ones. If tape is used, essentially all of these procedures involve duplication of the initial file in another position on the tape with corrections superimposed or added to the original SPSS systems file.

Erasure of taped data — If a SPSS systems file has been recorded on a tape, it will remain intact until erased. It is very easy to erase a tape by mistake. For example, suppose wet season data has been taped in position 1. The next set of data should be taped in position 2. If the programmer writes "1" on the JCL card where position should be noted, the dry season data will be taped over the wet season data — thereby erasing the first SPSS systems file. Consequently, all SPSS systems files should be duplicated as soon as they are constructed.

SAVE FILE for taped data — New variables, created through transforming the original variables (data modifications) may be saved in an updated systems file by use of the SAVE FILE procedure. Once the new variable is created, no record of the operations employed to construct the variable will be retained. For example, suppose you want to create an "input experience" index by summing the number of years each respondent has used fertilizer, insecticide, and several additional modern technologies. Once the index has been constructed and preserved through the SAVE FILE procedure, there will remain no record in the systems file of the variables used to construct the index. Consequently, it is critically important that the researcher makes a permanent record explaining how new variables were created. The DOCUMENT card permits the user to retain all information necessary to describe the data set, variables, etc. in the SPSS systems file and recall this information on any data processing run by means of the LIST FILE INFO card. The importance of fully annotating the systems file through the DOCUMENT card cannot be overemphasized. Unless this is done, it will be impossible for other researchers to use the systems file because they will be unable to determine the meaning or procedures used in constructing new variables that are not listed in the codebook.

This brief introduction to SPSS is intended to provide an overview as to what one must do to prepare data for computer analysis; and outline the type of analysis that is possible using a computer. In order to actually use SPSS, the researcher must carefully study the SPSS manual.

XIII. Conclusion

Data collected in a survey can make a valuable contribution to our understanding of the agricultural environment. Yet, the usefulness of the data collected dependent on the care taken in identifying our hypothesis, designing the questionnaire, conducting the interview, coding the data, executing the analysis, and writing the final report. Failure to follow the research principles applicable to each of these phases of the research project will greatly reduce the meaningfulness of the data collected.

Budget and time constraints may make it impossible to follow all of the guidelines outlined in this paper. Yet, to the extent possible, the researcher must attempt to carry out his research in accordance with the principle of survey research. Only when this is done, can the social scientist make a significant contribution towards increasing our understanding of the world in which we live.

REFERENCES

Babbie, Earl R. Survey Research Methods., Wadsworth Publishing Co., Belmont, Cal. 1973

Collinson. Farm Management in Peasant Agriculture, Praeger, New York, 1972.

CYMMIT. "Screening and Testing Farm Adopted Technologies: A Manual of Procedures for Agronomic-Economic Research", mimeo (draft), 1977.

DeDatta, S.K. et. A Handbook on the Methodology for An Integrated Experiment - Survey on Rice Yield Constraints, IRRI, 1978.

Hunt, K.E., Agricultural Statistics for Developing Countries, Oxford University, Institute of Agrarian Affairs, 1969.

Herdt, R.W. and R. Bernsten. "Methodology for Assessing Rice Yield Constraints". Working Paper 2, International Rice Research Institute, 1975.

Kearl, B. Field Data Collection in the Social Sciences: Experiences in Africa and the Middle East, Agricultural Development Council, New York, 1976.

Lynch, Fr. Frank. "How to Make a Social-Survey Interview Schedule: Instructions for Beginners". Institute of Philippine Culture, Anteneo de Manila University, Quezon City, undated mimeo.

Lipton, M.A. and M. Moore. "The Methodology of Village Studies in Less Developed Countries". IDS Discussion Paper No.10. Brighton, Institute of Development Studies, 1972.

Nie, N.H. et al., Statistical Package for the Social Sciences, 2nd Ed. McGraw Hill Inc., 1975.

Norman, D.W. "Problems Associated with the Gathering of Technical, Economic and Social Data at the Farm Level in West Africa". Paper presented at the International Conference on the Economic Development of Sahelian Countries, University of Montreal, Canada, October 13-14, 1977.

Parel, C.P. et al. "Sampling Design and Proceeures". Agricultural Development Council, New York, 1973.

Prewitt, Kenneth. Introduction to Social Science Methodology for East Africa. Publisher unknown, 1975.

Srinivis, M.N. "Village Living: A Source of Insight for Social Scientists". Agricultural Development Council Teaching Forum, Research Methodology (No.35, Jan. 1974).