

FAO
STATISTICAL
DEVELOPMENT
SERIES

2_a

MICROCOMPUTER-BASED
DATA PROCESSING
1990 WORLD CENSUS OF
AGRICULTURE



FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS

**FAO
STATISTICAL
DEVELOPMENT
SERIES**

2_a

**MICROCOMPUTER-BASED
DATA PROCESSING
1990 WORLD CENSUS OF
AGRICULTURE**

**FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS
Rome, 1987**

The designations employed and the presentation of material in this publication do not imply the expression of any opinion whatsoever on the part of the Food and Agriculture Organization of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

M-78
ISBN 92-5-102586-X

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior permission of the copyright owner. Applications for such permission, with a statement of the purpose and extent of the reproduction, should be addressed to the Director, Publications Division, Food and Agriculture Organization of the United Nations, Via delle Terme di Caracalla, 00100 Rome, Italy.

PREFACE

FAO has been promoting decennial world censuses of agriculture since 1950. The FAO Conference in its 23rd Session held in November 1985 approved the Programme for the 1990 World censuses and recommended that all member countries participate in the Programme by carrying out at least one national census of agriculture during the period 1986-1995.

It has been noticed that many developing countries face difficulties in the data processing phase of their censuses of agriculture; these difficulties cause long delays in the publication of the census results.

This booklet and a program diskette have been prepared by FAO to provide guidance on processing the census of agriculture data in an efficient way with minimal cost. Two versions of the program are available on request: interpreted or compiled. The interpreted version needs Base III to execute, while the compiled does not; the interpreted version can be modified according to users' needs, and the compiled cannot. A questionnaire designed to collect all census items that are proposed in the Programme for the 1990 World Census of Agriculture is the model for the input. A personal computer (IBM AT) and two software packages (dBASE III for the data entry and SAS for tabulation) have been used in developing the booklet and the diskette.

Deviations from the questionnaire will necessitate corresponding changes in the booklet and the diskette. Nevertheless, it is hoped that this booklet and the diskette will be helpful to plan and execute the data processing phase of a census of agriculture using a micro-computer.

Director
Statistics Division

CONTENTS

	<u>Page</u>
PREFACE	iii
1. INTRODUCTION	1
1.1 <u>Purpose</u>	1
1.2 <u>Readership</u>	1
1.3 <u>Programme for the 1990 WCA</u>	1
1.4 <u>Questionnaire</u>	2
2. CENSUS PROCESSING ON MICROCOMPUTERS	5
2.1 <u>Software</u>	6
2.1.1 Phases of the census data processing	6
2.1.1.1 Data entry	6
2.1.1.2 Editing	6
2.1.1.3 Tabulation	7
2.1.2 Integration of the data processing phases	7
2.2 <u>Hardware</u>	9
2.2.1 Minimum-equipped PC	13
2.2.2 Primary memory size	13
2.2.3 Secondary memory size	13
2.2.4 Monitor	13
2.2.5 Printer	14
2.2.6 Micro-to-mainframe link	14
3. INTEGRATED CENSUS AND SURVEY PROCESSING ENVIRONMENT USING dBASE III, SAS AND LOTUS 1-2-3	17
3.1 <u>Presentation of the software packages</u>	17
3.1.1 General information on dBASE III, SAS and LOTUS 1-2-3	18
3.1.2 Other tools used -- WORDSTAR and MICROSOFT WINDOWS	19
3.2 <u>Data entry system in dBASE III</u>	19
3.2.1 Characteristics and limitations of the model data entry system	20
3.2.2 Program logic and system files	21
3.2.3 Data entry organization and its support by data entry system	23
3.2.3.1 Monitoring the data entry by querying a log-file	24
3.2.3.2 Recovering from power-cuts	25
3.2.4 Variables	26
3.2.4.1 Naming conventions	26
3.2.4.2 List of variables	26
3.2.5 The data file structure	27
3.2.5.1 Combining data files	28
3.2.5.2 Creating dBASE III data files with the help of WORDSTAR	28

3.3	<u>Editing and correcting using dBASE III</u>	31
3.4	<u>Tabulation using SAS</u>	32
	3.4.1 Transferring the data from dBASE III to SAS	32
	3.4.2 Tabulation variables and their classes	33
	3.4.3 Generating tables in SAS	34
3.5	<u>Analysis and presentation using LOTUS 1-2-3 and SAS</u>	37
	3.5.1 Transferring tables from SAS to LOTUS 1-2-3	37
	3.5.2 Manipulation of tables in LOTUS 1-2-3	38
	3.5.3 Generating graphs in LOTUS 1-2-3	40
4.	<u>FUTURE DEVELOPMENT</u>	41
	4.1 <u>Software development</u>	41
	4.2 <u>Hardware development</u>	42
	APPENDIX A: Code of sample screen module and the MAIN module	43
	APPENDIX B: Listing of dBASE III file structures and variables	49
	APPENDIX C: Class descriptions for the main variables	55
	APPENDIX D: File structure of a sample file in SAS	65
	APPENDIX E: SAS data entry screens for PROC TABULATE	69
	APPENDIX F: Questionnaire Census of Agriculture 1990	71

CHAPTER 1

INTRODUCTION

1.1 Purpose

The purpose of this booklet is to provide guidance on micro-computer processing of census data to those countries that will participate in FAO's 1990 World Census of Agriculture (WCA) Programme. In the past, it has been observed that many developing countries faced difficulties in producing the final statistical tables from their censuses of agriculture following the completion of the field work. These difficulties were partly related to a lack of computer facilities and trained personnel. However, the emergence of microcomputers, which are much less costly than mainframe computers, and the availability of commercial software packages for microcomputers, have created new possibilities for processing agricultural census data. Microcomputers and the data processing software packages do not require a heavy investment on the part of users. This booklet is designed to assist those developing countries that wish to process the data collected in their censuses of agriculture with minimum expenditure on hardware and personnel training.

1.2 Readership

This booklet mainly addresses the technicians who are responsible for transferring the data from questionnaires into the microcomputer and for obtaining computer printed statistical tables. Accordingly, the booklet provides detailed lists of the operations involved in both data entry and tabulation.

It is assumed that the reader has a general understanding of the IBM-AT microcomputer (under MS-DOS) and the following software packages: dBASE III, SAS and LOTUS.

Although the basic thrust of this booklet is a discussion of the use of techniques of data processing, decision makers who would procure microcomputers for census data processing may find this booklet useful in evaluating the respective advantages of microcomputers and other computer systems.

1.3 Programme for the 1990 WCA

The census of agriculture for which data processing is described in this booklet, refers to a census taken along the lines explained in FAO's publication Programme for the 1990 World Census of Agriculture, Statistical Development Series, No. 2. The Programme for the 1990 WCA

is the latest of the world censuses of agriculture promoted by FAO and covers the 10-year period 1986-1995. The FAO Conference, at its 23rd Session held in Rome during November 1985, approved the Programme for the 1990 WCA and urged all countries to participate in this programme by conducting a census of agriculture during the 10-year period mentioned above.

The Programme for the 1990 WCA presents the census of agriculture as a component of an integrated national statistical programme on food and agriculture. It is recommended that the census of agriculture not be overburdened with too many items: the census of agriculture should cover only those items related to agricultural structure, and data on other items of interest should be collected through specific surveys or other sources of data. Based on the experiences of previous rounds of the WCA, the Programme for the 1990 WCA proposes a selected list of census items presented in Chapter 3 of the Programme for the 1990 World Census of Agriculture.

1.4 Questionnaire

In the preparation of this booklet it was assumed that the questionnaire given in Appendix F has been used in collecting the data. This questionnaire includes all census items proposed in the Programme for the 1990 WCA and has been redesigned after field tests in Chile (1985) and Turkey (1986). The response cells for most of the questions are precoded.

The questionnaire opens with a section titled "IDENTIFICATION OF THE QUESTIONNAIRE". This section contains information about the location of the holding: Province, District, Enumeration area and Serial number of the questionnaire. It includes space for "expansion factor" which is the inverse of the probability that the holding being interviewed has been selected. For complete enumeration, the expansion factor is 1.00 for all holdings. For sample census enumeration, the expansion factor must be calculated and preferably recorded on the questionnaire during the editing phase.

The rest of the questionnaire consists of nine sections, each of which covers a particular set of related items. The sequence of the "parts" follows the order adopted in the Programme for the 1990 WCA. It should be noted that any of the items not applicable to a given country should be eliminated from the questionnaire. Conversely, some countries may wish to add new items to the questionnaire. Before adding new items, countries are recommended to consult the Programme for the 1990 WCA, paragraphs 2.7-2.10, where general recommendations are provided for the scope of the census. Deletions from and additions to the questionnaire given in Appendix F would require changes in the computer programs provided in this booklet.

Part 1 of the questionnaire is "IDENTIFICATION OF HOLDER, RESPONDENT, HIRED MANAGER". Some of this information is for control and subsequent follow-up activities and as such they need not be entered into the computer. Part 2 is on "GENERAL CHARACTERISTICS".

Part 3 covers "DEMOGRAPHIC CHARACTERISTICS AND EMPLOYMENT". Section A records data on the holder and his/her household members; Section B records data on hired workers. The first part of Section A is devoted to the holder. In question 3.7 demographic and employment data are collected from members of the holder's household excluding the holder. Employment data refer only to those members who are older than 11 years. Two codes are assigned to each member older than 11 years; for those who worked, the code refers to the number of months worked "outside the holding" and the "total" months worked during the preceding 12 months; for those who did not work, the code indicates whether or not the individual looked for a job. Two codes are indicated in the last two columns of question 3.7: The first code identifies whether the member is economically active or not. Countries must decide the total number of months worked that would justify inclusion of the members, particularly homemakers and students, in "economically active". Reference should be made on this point to paragraphs 3.40-3.46 of the Programme for the 1990 WCA. The second code is for classifying the members as "permanent worker on the holding", "occasional worker on the holding" or "did not work on the holding". Countries again must decide on the minimum number of months worked on the holding to distinguish "permanent" workers from "occasional" workers. The Programme for the 1990 WCA provides more information on this subject in paragraph 3.54. These two codes could be recorded in the questionnaire during the editing phase. Data on each household member is entered into a computer file that reserves space for 10 members.

Part 4 of the questionnaire details "LAND AND WATER". Questions 4.6 and 4.7 require collection of data for each parcel in addition to requesting holding level totals. The main reason that data are collected at the parcel level is to improve the quality of data at the holding level. If questions are directly asked at the holding level, calculations may be made mentally or on scratch paper and may result in inaccurate totals. It is expected that holding level totals will be checked during the editing phase of the questionnaires. Only holding level totals are entered in the computer.

In part 5 of the questionnaire, "CROPS" are covered. For questions 5.5, 5.6, and 5.7, where data are recorded at the parcel level and then at the holding level, the explanation given in the above paragraph applies. Countries should prepare a code list for crops grown and should decide which crops to include in the questionnaire (if any) as precoded.

Parts 6 and 7 of the questionnaire address "LIVESTOCK" and "MACHINERY and EQUIPMENT", respectively. Countries should prepare a code list for livestock types and a code list for agricultural machinery. They may opt to include as precoded some of the livestock types and machinery in the questionnaire.

Part 8 of the questionnaire provides data about "NON-RESIDENTIAL BUILDINGS". Question 8.2 requires that data be collected for each building separately. However, only the holding level totals are recorded in the computer. These totals need to be checked during the editing phase of the questionnaires. In the code column under "Tenure" of question 8.2 the following codes are proposed:

- 1 for owned;
- 2 for rented; and
- 3 for other.

Part 9 is the last part of the questionnaire and covers "ANCILLARY ACTIVITIES"; specifically, forestry and fisheries.

CHAPTER 2

CENSUS PROCESSING ON MICROCOMPUTERS

This booklet presents an integrated environment for census and survey processing based on the standard software packages dBASE III, SAS, and LOTUS 1-2-3. An important issue detailed below is the data transfer between packages; the integrated environment is open-ended and new packages can be easily added. The environment presented is modular and flexible.

As noted above, the objective of this booklet is to provide guidance in the field of data processing for the 1990 WCA. Although the booklet primarily addresses the use of microcomputers (PCs), a majority of software packages discussed in this booklet, such as SAS, CENT and CONCOR, have long been available on mainframes. One can soon expect a proliferation of integrated environments consisting of mainframe computers with a number of PCs attached to them.

The booklet can be used to establish strategies for the processing of future censuses and surveys. Future solutions can be planned in terms of hardware and software, on the basis of the size of expected data, the number of workstations needed to enter the data, etc. It should be underlined that development of microcomputer technology will bring less expensive PCs, larger secondary storage and multi-user operating systems, all of which will help make more feasible census processing with microcomputers.

The number of questionnaires that can be processed on microcomputers will most likely be of interest to the reader. This is a difficult question to answer, and the answer may change from month to month. On the latest PC -- COMPAQ 386 computer -- 130 MB secondary storage is available; this feature offers the possibility of storing 130,000 to 260,000 questionnaires. Additional storage can be purchased. A bottleneck in census and survey processing can occur when there are too few data entry stations. This condition can be improved by increasing the number of data entry shifts.

Future trends that will influence the size of statistical tasks manageable with PCs are discussed in the chapter entitled "Future Development". Assuming a very rapid development in the field of microcomputers, one must expect that the design of microcomputer-based statistical environment will change considerably in the near future.

2.1 Software

2.1.1 Phases of the census data processing

The processing of a census or survey involves these tasks:

data entry (from questionnaires);
editing (correction, imputation);
tabulation;
analysis; and
presentation.

2.1.1.1 Data entry

The most versatile software packages are described below:

The United States Department of Commerce, Bureau of the Census, has compiled a list of four leading data entry packages. After preliminary evaluation, the U. S. Bureau of the Census has focused on two packages that may be used for data entry. The first package, ENTRYPOINT is programmer oriented; the second one, RODE/PC is user-friendly. Both packages have their advantages and are targeted toward somewhat different environments.

dBASE III can be used for data entry but error-checking is very rudimentary and extensive programming is needed if errors such as consistency errors are to be detected at the data entry phase.

The data entry task requires a large number of data entry stations (terminals) but poses a low demand on data processing capability. Therefore, one should seek a method for inexpensive expansion of the AT that utilizes several terminals (or older 8-bit PCs emulating terminals). One possible solution is the installation of the multi-user operating system XENIX on AT. A multi-user version of MS-DOS is expected soon.

2.1.1.2 Editing

The term "editing" is frequently used to describe the process of "cleaning" the entered data. The process includes checking all data for possible value errors and consistency errors. Inconsistent values must be corrected either by consulting the related questionnaire or by automatic imputation. Frequently, the error is in the questionnaire and the only solution is an imputation.

It is the authors' conviction that, if possible, the data should be cleaned at the data entry stage. This is a relatively advantageous strategy with a limited amount of data, which is the case for surveys.

Very few data correction software packages are available. The mainframe version of CONCOR has been used for a long time for census processing. The US Bureau of the Census has transferred the CONCOR package to PCs; this package can be obtained from the US Bureau of the Census, but to run the program, a COBOL compiler is necessary.

2.1.1.3 Tabulation

Several tabulation packages are available for the tabulation of survey and census data.

CENT, COCENT, XTALLY are (were) the packages widely used for the census tabulation. Some of these packages are now available on PCs. One can obtain CENT 4 from the US Bureau of the Census. However, the package requires a COBOL compiler. The price of the Realia COBOL compiler (recommended by the US Bureau of the Census) is US \$ 1 000.

Alternatively, several commercial statistical software packages can be used for the tabulation; the best known are SAS and SPSS.

2.1.2 Integration of the data processing phases

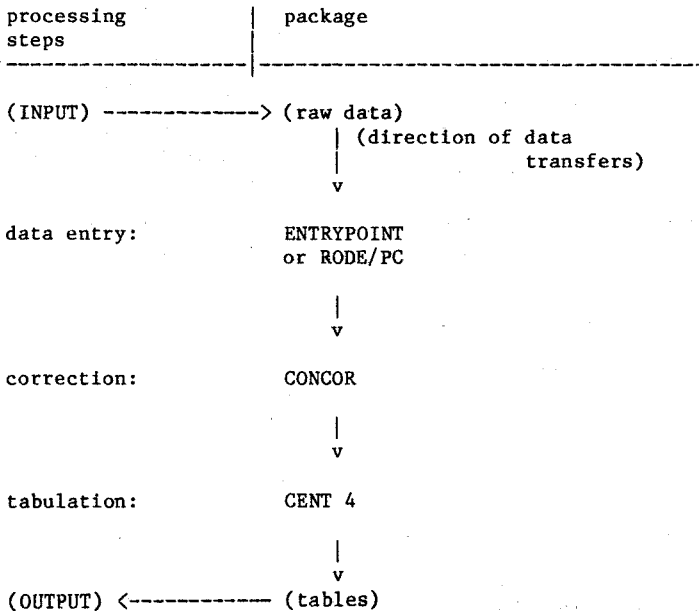
For every phase of the census processing task, two solutions are possible:

- a. custom-developed programs; and
- b. software packages.

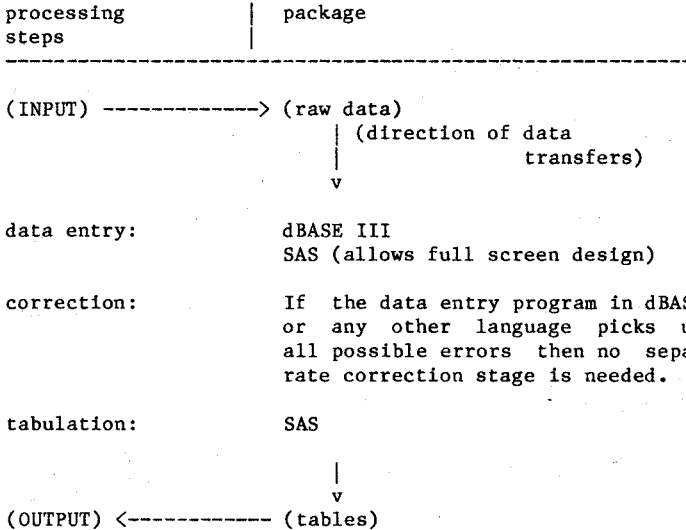
Various types of software may be combined into integrated processing chains. The major problem in such an effort is to ensure compatibility of data files, so that output data of one program can be directly used by other programs.

Several survey and census processing software chains are possible;

Processing Chain 1



Processing Chain 2



In the above chain SAS can be replaced by SPSS. Several other combinations and processing chains are possible. However, care should be taken to ensure a minimum of data conversions between various packages.

Described below is Data Processing Chain 2. The chain will be extended during the analysis stage. For this purpose, the SAS package is an excellent alternative. Presentation of the tables and an additional restructuring of the census tables can be accomplished using the LOTUS 1-2-3 spreadsheet. LOTUS 1-2-3 can also be used to produce a graphic presentation of the data.

The detailed information on the GENT 4 and CONCOR packages needed to implement Data Processing Chain 2, may be obtained from:

International Statistical Programs Center
Bureau of the Census
United States Department of Commerce
Washington, D.C. 20233
USA

2.2 Hardware

The focus of this booklet is microcomputers rather than larger computers. At present, the most popular microcomputers (PCs) are various IBM PC models and compatibles. Those PCs termed "IBM compatible" are produced by manufacturers other than IBM, but follow a de facto "standard" of IBM PC architecture. There are various degrees of compatibility with IBM PCs, but basically the compatibles should be able to run the same software as an IBM PC.

The IBM PC base system (PCG) normally includes 64 kb memory on the motherboard and one 360 kb diskette drive. A PCG equipped with a hard disk and extra memory is basically the same computer as the XT. Due to primary memory and secondary storage requirements of modern software, it is advisable to acquire an XT. Each software package requires a minimum amount of primary memory (RAM) but an application will run more efficiently if more memory is available. Thus, 512 kb RAM should be a reasonably good primary memory size, although this may change with the introduction of a multi-user operating system and more memory may be needed. The basic PCG may be enough for word processing and data entry or can be used as a node in a network of PCs. For most processing tasks either an XT or AT or compatible should be chosen.

The latest innovation in the PC family is the 80386 processor. COMPAQ introduced the PC/386 in November 1986; it is AT compatible. The PC/386 contains a 130 MB hard disk and its price is only slightly higher than that of the IBM AT. It is expected that 80386-based computers also will be available from other manufacturers.

Table 1. Comparison between PC AT and PC XT

Model	AT	XT
processor	Intel 80286	Intel 8088
co-processor	* Intel 80287	* Intel 8087
max. primary memory (RAM) on the motherboard	512 kB	256 kB
expansion by plug-in board	* 128 kB	* 384 kB
secondary memory (hard disk)	20 MB	10 MB
additional hard disk installation	* 20 MB within the AT	* 10 MB PC Expansion Unit needed
diskette drive	1,2 MB	360 kB (full size)
additional diskette drive	* 1,2 MB or 360 kB (half-height size)	* 360 kB (half-height) (2 half-height drives)
parallel port	included	on a monitor adapter
serial port	included	included
clock	included	* optional

* signifies that the item is optional

Table 1. Comparison between PC AT and PC XT (cont)

monitors (one to be selected)	-monochrome display -color graphic display -Enhanced Color Display	-monochrome display -color graphic display -Enhanced Color Display
monitor support cards (one must be selected and must be compatible with the monitor)	-monochrome adapter -color graphic adapter (640 x 200 pixels) -Enhanced Graphic Adapter (640 x 350 pixels) Other cards, e.g., Hercules	-monochrome adapter -color graphic adapter (640 x 200 pixels) -Enhanced Graphic Adapter (640 x 350 pixels) Other cards, e.g., Hercules
printers	parallel or serial IBM PROPRINTER and EPSON printers are most popular	parallel or serial IBM PROPRINTER and EPSON printers are most popular
standard configuration	512 kB 20 MB hard disk 1,2 MB diskette drive	256 kB 10 MB hard disk 360 kB diskette drive
recommended additions	360 kB diskette drive (to ensure compatibility with XT) Enhanced Graphic Adapter (supports all three monitors -monochrome -color graphic -Enhanced color)	256 kB memory expansion card (can be combined with a serial port and the clock) Enhanced Graphic Adapter (supports all three monitors -monochrome -color graphic -Enhanced color)
recommended configuration	-512 kB RAM -20 MB hard disk -1,2 MB diskette drive -360 kB diskette drive -Enhanced Graphic Adapter -Enhanced Color Display or Monochrome Display -printer (matrix) (special quad density diskettes are needed for 1,2 MB drive)	-512 RAM -multifunctional card with 256 kb RAM,(serial port) and clock -10 MB hard disk -one 360 kB diskette drive or two half-height 360 kB diskette drives -Enhanced Graphic Adapter -Enhanced Color Display or Monochrome Display -printer (matrix)
estimated cost in US \$ in USA	6 000 - 8 000	4 000 - 5 000

Table 2. Preliminary data on the latest generation of PCs

Model	COMPAQ 386
processor	Intel 80386
secondary memory (hard disk)	130 MB
diskette drive	1,2 MB
parallel port	included
serial port	included
clock	included
estimated cost in US \$ in USA	8 000

minimum-equipped PC (integrated workstation, data entry station)

recommended configuration (LAN, PC-mainframe connections)

- 512 kB RAM
- one diskette drive
- monochrome monitor

optional

- PC-Mainframe connection board
- LAN connection board
- serial interface card

2.2.1 Minimum-equipped PC

The processing of census data on PCs can be adversely affected by an insufficient number of data entry stations. To dedicate an AT for data entry may be a waste of resources; a minimally-configured PC should be used for data entry. Such a PC should contain a monochrome monitor, 256/512 kb main memory, and one 360 kB diskette drive. The cost of such a minimally-configured PC will be about US \$ 1 000-2 000, which is somewhat less than the price of an OEM asynchronous terminal. At the upper range one can purchase an AT without hard disk so that, later on, additional storage can be added.

2.2.2 Primary memory size

256 kB is the minimum memory needed to run a majority of software packages although some software may require 512 kB. It is an advantage to have as much primary memory (RAM) as possible. An upper limit of 640 kB memory can be accessed by the present version of the MS-DOS operating system. For additional information about memory requirements, one should consult the software manual.

2.2.3 Secondary memory size

Two diskette drives is a standard configuration if no hard disk is installed. However, several packages, including SAS and SPSS, require a hard disk. Census and survey processing will require relatively large secondary memory in order to store all the data. While add-on hard disks are available in various sizes from 10 to 250 MB, it is advisable to check whether the software will run on the add-on disks -- especially on the larger disks of 30 MB and upwards -- before purchasing an additional disk.

2.2.4 Monitor

A majority of spreadsheet packages require graphic capability; either a Color Graphic Adapter/Color Monitor or an Enhanced Graphic Adapter/Enhanced Color Display should be purchased. The Hercules card also supports graphics on a monochrome monitor. Graphics and color are usually implemented together but the color is rarely a necessity. The Color Graphic Adapter (640 x 200 points) is not suitable for word processing because the resolution is poor; the Enhanced Graphic Adapter (EGA) is preferred. Some software does not yet support the EGA card in the high resolution mode. The card can be operated in both the (640x200) and (640x350) mode. The standard IBM monochrome monitor is of poor quality and should be, if possible, replaced by another compatible monochrome monitor.

2.2.5 Printer

There are two kinds of communication interfaces between a computer and the printer: parallel (most common) and serial. The parallel interface is provided on a PC as "standard". Most of the packages support 10 to 20 of the most popular printers. The graphic output is usually dependent on the graphic capability of the printer. The number of columns the printer can print may be important. Printers designed

to handle 80 or 132 columns in the 10 characters per inch mode are available; for the purpose of printing statistical tables, printers capable of printing 132 characters will be needed. Most of the printers are also capable of printing 12 and 17 characters per inch.

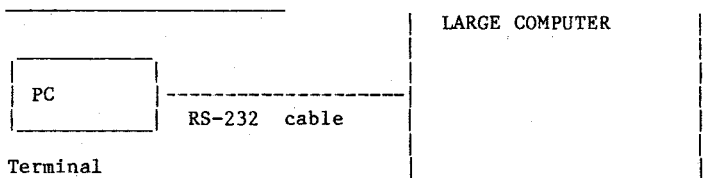
2.2.6 Micro-to-mainframe link

Although the environment discussed in this booklet is geared toward microcomputers, there is nothing that could prevent an incorporation of larger computers into the configuration. The SAS package and several census packages are available on mainframes.

There are two possible solutions for the communication between a terminal and a CPU -- one is defined by the RS-232/V24 standard and the other by various synchronous protocols. The most popular synchronous protocol is the IBM protocol.

The discussion has thus far referred to communication between a terminal and a CPU without mention of a PC. From the point of view of a mainframe, a PC is seen as a terminal. Thus, most terminals can be replaced by PCs and new PCs can be added instead of terminals.

The majority of minicomputers communicates with terminals using the RS-232 standard. To establish communication between a computer equipped with an asynchronous port and a PC one will need an asynchronous (serial) adapter board in the PC. In fact the AT has as standard such a serial connection. A PC and the computer should be connected with an appropriate cable. In addition, a PC will need a terminal emulation software package in order to "play" the role of a terminal. Several communication software packages that emulate a terminal are available on PCs.



Terminal emulation software; for example, KERMIT

If data transfer of binary files is planned, the same communication software used on the PC must be installed on the mainframe

For many purposes one would like to transfer data from a PC to a larger computer or from a larger computer to a PC. The additional capacity to transfer not only text files but also binary files, will require that a special software package be used on the larger computer. At present, KERMIT protocol is the most widely used protocol.

Because IBM mainframes and plug-in compatibles are so widely used, the discussion below of synchronous communication will concentrate on 3270-type communication. In order to communicate with an IBM mainframe or compatible a special card must be installed inside a PC. Several of these products are on the market. The cards or boards are relatively expensive (US \$ 1 000 - 1 500). The boards emulate an 3270-type terminal on a PC and in addition allow transfer of data between a mainframe and a PC.

Type of communication	PC end		Mainframe end	
	hardware	software	hardware	software
serial, RS-232 terminal emulation data transfer	Asynchronous serial board	KERMIT CROSSTALK KERMIT BLAST	Asynchronous port (usually standard)	KERMIT BLAST
synchronous (3270-type) terminal emulation and data transfer	IRMA board PCOX board FORTE board	(included) (included) (included)	(items already in operation, unplug terminal -- plug in PC) 3270-port (cluster controller)	TP monitor (drives terminals)

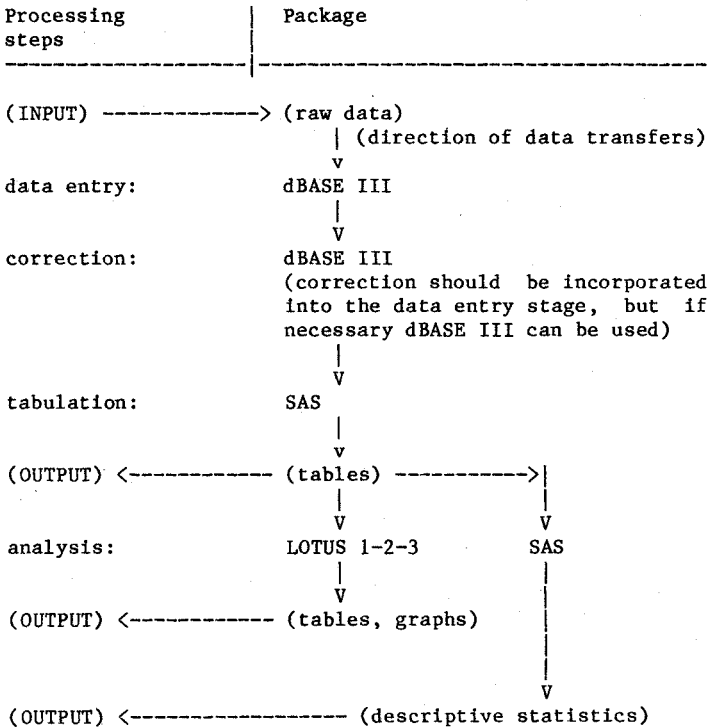
Several approaches are possible for a micro-mainframe integrated environment. One may enter data on PCs, then move the data to a mainframe for tabulation. If there is an older minicomputer with several terminals but without tabulation software, then the data can be entered on a mini and moved to a powerful PC for tabulation. SAS supports data transfer, in SAS dataset formats, between PCs and a mainframe using IRMA or PCOX cards.

CHAPTER 3

INTEGRATED CENSUS AND SURVEY PROCESSING ENVIRONMENT USING
dBASE III, SAS and LOTUS 1-2-3

3.1 Presentation of the software packages

The integrated environment described in this booklet is illustrated in the following schema.



In order to run an integrated microcomputer-based environment the following minimum hardware configuration is assumed:

- IBM or IBM-compatible PC
- 512 kb RAM
- at least 10 MB hard disk
- monitor
- matrix printer

3.1.1 General information on dBASE III, SAS, LOTUS 1-2-3

dBASE III/dBASE III Plus

dBASE III is the most popular data base software on microcomputers. It is estimated that dBASE III accounts for 85 percent of all data bases installed on microcomputers.

There is only a slight difference between dBASE III and dBASE III Plus, from the point of view of a single user. dBASE III, with a few additional commands, also runs in a network environment.

The dBASE III data entry program described in this booklet and listed in Appendix A, can be obtained from FAO. In order to run the program, the dBASE III software package is necessary. Most hardware vendors will be able to supply this software.

dBASE III has a very simple internal structure: a file is equal to a relation; a relation is a set of rows (tuples); one row can be seen as an observation (case) -- a row is also termed a record in computer terminology; each row (tuple, record) is divided into fields; a field is part of a column going across all rows; a column describes a single attribute (property, characteristic), and is also a variable. Thus, a field stores a value of a particular variable for a particular observation. The terms in parentheses express various terminological schemes used by various "schools". Relational algebra, computer science, statistics, and even SAS itself all have their own terminological preferences.

SAS

SAS stands for Statistical Analysis System. The package has developed from a purely statistical software toward almost a fourth generation language for general use. SAS is available on the following microcomputers: IBM PC, PC/XT and IBM AT under PC DOS. SAS is available on the following mainframes and minicomputers: IBM 370, 30xx, 43xx and compatibles; on Digital Equipment Corporation VAX 8600 and the 11/7xx series under VMS, microVAX; the Data General ECLIPSE MV series under AOS/VS; and Prime Computer.

dBASE III runs only on microcomputers, but SAS is available on larger computers. Data could be entered on microcomputers (in dBase or otherwise) and subsequently transferred to a larger computer for processing with SAS.

LOTUS 1-2-3

LOTUS 1-2-3, Version 2 is a spreadsheet package with some graphic functions. Lotus version 2. LOTUS allows the import and export of data in several data exchange formats: DIF, DBF, and SYLK; this allows the

integration of LOTUS with other software packages. LOTUS provides a macro possibility. When properly written, macros may allow an "automatized" manipulation of a series of spreadsheets with a similar structure. The macro feature is a help in manipulating sets of statistical data with a uniform structure.

3.1.2 Other tools used - WORDSTAR and MICROSOFT WINDOWS

The WORDSTAR word processor/editor was extensively used throughout all stages of the system preparation and data manipulation. WORDSTAR was used to:

1. Write dBASE III programs.
2. Create hundreds of variable names. This text information was then submitted to dBASE III using CREATE <DBF-filename> FROM <TXT-filename>, where TXT-filename is a file structure created by WORDSTAR.
3. Delete unnecessary table construction elements before moving the data to LOTUS 1-2-3. Likewise, the left-justified stubs were prefaced by a single quotation mark before transfer to LOTUS.
4. Add comments and footnotes.

A word processor/editor preferably should have the following functions in order to assist in manipulating data:

1. An editor function so that pure ASCII files can be produced without embedded formatting information;
2. A CHANGE/REPLACE function for all ASCII characters, even LINE FEED and CARRIAGE RETURN characters;
3. BLOCK COPY and MOVE commands so that any piece of text can be moved to another place. For example, the editor should be able to move a box defined by "upper left corner LINE 5 COLUMN 10 and lower right corner LINE 20 COLUMN 40." Basically, this option allows the movement of table columns and insertion of new columns in a table; and
4. The ability to group commands in MACROs so that repetitive sequences can be easily performed.

WORDSTAR does not provide a MACRO facility.

MICROSOFT WINDOWS was used to make snapshots of SAS menus. SAS can be run within WINDOWS but at present cannot be "iconized" so every time data are moved to a CLIPBOARD, one must exit SAS in order to save the content of the CLIPBOARD.

3.2 Data entry system in dBASE III

The process of developing data entry software can be seen as an exercise that tests the proposed data processing environment, checks the interface capacity of various packages and, as a by-product also tests the proposed design of the 1990 WCA questionnaire.

The data entry system described in this booklet is based on the questionnaire prepared for the 1990 WCA by the Statistics Division of FAO. The data entry system is a model system that should be modified according to the specific design of an actual questionnaire.

The system is written in dBASE III, the most popular data base system for microcomputers. While the data entry system could have been written in any computer language, the choice of dBASE III is based on its popularity and flexibility. Most of the dBASE III code is self-explanatory and it should be easy to understand and modify. DBASE III produces the data in several formats that make data transfer simple. Several packages including SAS will accept dBASE III formats.

The model data entry system is available in two versions, interpreted and compiled. The compiled version does not need dBASE III software for execution as it is a pure demonstration version. The interpreted version contains the dBASE III source code and can be altered for particular user needs.

The booklet shows how several packages can be integrated with each other. The data is entered using dBASE III in one of two ways. The simple method uses the dBASE III ASSIST option, but the controls are only rudimentary. To check for consistency errors, it is necessary to put the commands of dBASE III in sequence, which is very similar to traditional programming.

Alternatively, if a more elaborate check routine must be implemented then the data entry system written in dBASE III is available. The "model" data entry system, described in the booklet, can be obtained from FAO. The "model" data entry system must be modified to reflect the particular conditions of the country.

3.2.1 Characteristics and limitations of the model data entry system

As noted above, the data entry system is a model data entry system that should be modified according to requirements of an actual questionnaire. In addition to structural changes that might be needed, the following characteristics might need modification:

1. The identification code of a questionnaire consists of 4 fields, each 2 characters wide.
2. Some of the questions may have a long list of variables associated with them. Thus, a household may contain a number of persons and each person is described by several characteristics, each characteristic being a variable. Also, each crop cultivated on a holding will be a variable. But the total number of possible crops (in the country) is limited. In the model data entry system several questions permit at present only groups of five variables. For example, the model system handles only five crops, five livestock types, etc. Additional variables should be created and the upper limits of the loops increased. No changes in the logic of the system are necessary.

3. The system consists of 18 screens, each corresponding to a part of the questionnaire. These screens can be compounded so more information is requested from one physical screen. Current screens can be deleted or new screens can be inserted.
4. Lead texts to the questions consist only of question number; full text could be incorporated.
5. The system checks for range errors and logical consistency errors. Additional checks may be needed.
6. Some checks of upper and lower limits of non-logical variables, (for example, areas) should be implemented to accommodate conditions of a particular country.

Once the user modifies the data entry system and tests it using the dBASE III system, the user may compile the dBASE code, as he compiles any other code, for example COBOL or FORTRAN. To compile the code, a dBASE compiler is needed. CLIPPER is one of the dBASE III compilers available on the market. The compiled code can be executed on several PCs without the dBASE III system being installed on the machines.

3.2.2 Program logic and system files

The full system consists of the following files:

C-E.PRG
C-ENTX.PRG
C-CODE.DBF
C-LOG.DBF
C-1.DBF
C-2.DBF
C-3.DBF
C-4.DBF

The data entry system is contained on 2 program files,

C-E.PRG (start up file)
C-ENTX.PRG (main program and all data entry screens)

The system uses four files on which data is stored. These are:

Database Files	# Records	Last Update	Size
C-1.DBF	0	86.08.29	3072
C-2.DBF	0	86.08.28	512
C-3.DBF	2	86.08.29	3072
C-4.DBF	2	86.08.29	3308

Two auxiliary system files are used by the system:

Database Files	# Records	Last Update	Size
C-CODE.DBF	2	86.08.28	302
C-LOG.DBF	2	86.08.29	528

Both files are used by the system to control access to the system and to record the progress of the data entry. The system monitors the entered questionnaires thus providing a logistic control that ensures that the questionnaires are entered in sequence.

The structure of these auxiliary files is as follows:

- . USE C-CODE
- . DISP STRU

Structure for database: C:C-CODE.dbf

Number of data records: 2
Date of last update : 86.08.28

Field	Field Name	Type	Width	Dec	Comments
1	CODE	Character	3		personal code
2	NM1	Character	15		full name
3	NM2	Character	15		second name
4	F1	Numeric	10		can be used as required
5	F2	Numeric	10		
** Total **			54		

- . USE C-LOG
- . DISP STRU

Structure for database: C:C-LOG.dbf

Number of data records: 2
Date of last update : 86.08.29

Field	Field Name	Type	Width	Dec	Comments
1	CODE	Character	3		personal code
2	OK	Character	1		structural check
3	QO_1_S	Numeric	2		id 1 start
4	QO_2_S	Numeric	2		id 2 start
5	QO_3_S	Numeric	2		id 3 start
6	QO_4_S	Numeric	2		serial start
7	QO_4_E	Numeric	2		serial end
8	AMOUNT	Numeric	2		number quest. entered
9	DATE	Date	8		data entry date
10	TIME_S	Character	8		time start
11	TIME_E	Character	8		time end
** Total **			41		

C-CODE.DBF contains the personal codes for the data entry staff. When starting the system, the user is asked to provide a code. This code is compared with codes in the C-CODE file. Only if the code entered is successfully matched by a code in the C-CODE file, is the user allowed to enter the data. The supervisor is responsible for addition or deletion of codes from the C-CODE file.

C-LOG.DBF contains a log for each computer session. The file contains information on the starting identification codes for the first questionnaire entered during the session. After the session has ended, identification codes for the last questionnaire entered during the same session are recorded. The file also contains starting and ending times for the session.

If a session ends in a controlled manner (no power cuts or other unexpected events) then a log-record has a sign "Y" in the field OK, otherwise, the field will contain the value "N". If the previous session ended in an uncontrolled manner, then the fields Q0 1 E through Q0 4 E will contain an identification for the last successfully entered record. It is the supervisor's task to restore the consistency of the data base by checking, with the help of dBASE III, the last records on each of the files C-1, C-2, C-3, and C-4.

The system will not allow the entry of any new data as long as there is inconsistency in the data base.

The data entry software (C-ENTX.PRG file) contains 18 modules. Each module corresponds to a part of the questionnaire and all modules are built in the same way. Each module is a procedure that is accessed from the kernel of the system. Module no. 2 has been selected to illustrate the process, and additional comments are also provided. Module no. 2 is listed in Appendix A; other modules have a similar structure.

3.2.3 Data entry organization and its support by data entry system

One of the most problematic aspect of data entry is the organization of work. Organization includes: management of the data entry personnel, physical organization of the questionnaires and establishment of routines for distributing questionnaires to the typists. Frequently, the data entry is done by staff without previous experience in such an undertaking. In some developing countries, it is easier to hire a computer expert to manage the processing aspect than to establish a rigorous organizational structure for data entry.

The model data entry system described in this booklet monitors the data entry and alerts the manager if there is something wrong in the organization of entered questionnaires.

The data entry system performs several managerial tasks:

- (1) monitors who are entering data;
- (2) records an identification of the first and the last questionnaires entered;
- (3) provides information about the start- and end-time of the data entry session;
- (4) provides the number of questionnaires entered;
- (5) prevents entering of questionnaires out of sequence;
- (6) prevents double entry; and
- (7) salvages the entered data, with the exception of the last record, in case of a power cut.

3.2.3.1 Monitoring the data entry by querying a log-file

From an organizational point of view, the following are the prerequisites for data entry using the system described:

- (1) Coded questionnaires must be given clear and unique identification numbers, i.e., province, district, enumeration area and serial numbers; and
- (2) A subject matter specialist should be available at a computer site to resolve inconsistencies in the data.

C-LOG contains the information about the progress of the data entry.

```
code OK pro dis enu beg end # date      time_s  time_e
AAA Y  99 99 99  1  4  4 86.09.01 08:54:07 08:55:14
AAA Y  88 88 88  1  3  3 86.09.01 08:55:15 08:55:44
AAA Y  99 99 99  5  7  3 86.09.01 08:57:06 08:57:39
```

where

pro is the province in our "model" country
dis is the district
< enu is the enumeration area
beg is the starting serial number of record entered
end is the serial number of the last record entered
is the number of questionnaires entered
date is in the format (year/month/day)
time_s is the starting time
time_e is the ending time.

A record on the C-LOG file is created after each data entry session or for each enumeration area. Thus, the system may create several records on a log-file during one session if more than one enumeration area is entered during one session.

For example, to find the number of questionnaires entered by a key-in operator with code "AAA" the following command can be used:

```
. SUM AMOUNT TO M_SUM FOR CODE = "AAA"  
. ? M_SUM  
10
```

To find the total number of questionnaires for province 99 the command is:

```
. SUM AMOUNT TO M_SUM FOR QO_1_S = 99  
. ? M_SUM  
7
```

3.2.3.2 Recovering from power-cuts

A power cut during a data entry will most probably result in a disrupted data structure. The last record entered will be lost. Actually, the situation is a bit more complicated. As one questionnaire contributes to several files, the questionnaire may result in records in files C-1 and C-2 but not in file C-3. This condition would occur because the power cut came before the record was stored in file C-3. The following log file shows that something has happened following the entry of questionnaire 88 88 88 5 as the last record has a "N" character in the field OK, instead of a "Y" as in all previous records. The system will stop working until consistency of the structure is re-established.

code	OK	pro	dis	enu	beg	end	#	date	time_s	time_e
AAA	Y	99	99	99	1	4	4	86.09.01	08:54:07	08:55:14
AAA	Y	88	88	88	1	3	3	86.09.01	08:55:15	08:55:44
AAA	Y	99	99	99	5	7	3	86.09.01	08:57:06	08:57:39
AAA	N	88	88	88	4	5	2	86.09.01	09:05:18	09:05:32

At this stage, the supervisor should go through all data files C-1, C-2, C-3 and C-4 and check the last records on each file. The interrupt occurred during the entry of questionnaire no. 6. The last consistent questionnaire is questionnaire no. 5. Because some of the data files may contain record no. 6, this record should be deleted. Once all data files contain record no. 5 as the last entered record the supervisor may proceed with the next step.

```
. GO 4  
. REPLACE OK WITH "Y"
```

code	OK	pro	dis	enu	beg	end	#	date	time_s	time_e
AAA	Y	99	99	99	1	4	4	86.09.01	08:54:07	08:55:14
AAA	Y	88	88	88	1	3	3	86.09.01	08:55:15	08:55:44
AAA	Y	99	99	99	5	7	3	86.09.01	08:57:06	08:57:39
AAA	Y	88	88	88	4	5	2	86.09.01	09:05:18	09:05:32

Now the key-in operator will be able to continue to enter the data.

3.2.4 Variables

3.2.4.1 Naming conventions

Variables are given names according to the following schema:

Qa_b_ccd

where

Q is always present and every variable starts with Q.

"a" and "b" point to the question, thus Q3_1_ccd refers to question 3.1 in the questionnaire.

"cc" has value 00 for a simple question. When the question results in a matrix of cells, "cc" refers to a row and "d" to a column. "cc" takes numbers as reference and "d" letters. Thus, variable Q3_1_12C refers to a variable in row 12 column C of the question numbered 3.1.

3.2.4.2 List of variables

Each question of the questionnaire requires one or more variables to store the data.

A listing of a few variables of the file C-1 follows. The full list of variables of the model data entry system is provided in Appendix B.

- . USE C-1
- . DISP STRU

Structure for database: C:C-1.dbf

Number of data records: 1

Date of last update : 08/11/86

Field	Field Name	Type	Width	Dec
1	Q0_1	Numeric	2	
2	Q0_2	Numeric	2	
3	Q0_3	Numeric	2	
4	Q0_4	Numeric	2	
5	Q1_2_00Z	Numeric	1	
6	Q1_4_00Z	Numeric	1	
7	Q1_7_00Z	Numeric	1	
8	Q1_9_00Z	Numeric	1	
9	Q2_1_00Z	Numeric	1	
10	Q2_2_00Z	Numeric	1	

3.2.5 The data file structure

The system uses four files, C-1, C-2, C-3, and C-4 because dBASE III can only accept 128 fields per file, though several files can be opened simultaneously. dBASE III stores data in relations. Each tuple is an object, case, or observation, while columns are variables, attributes, or properties.

Files C-1, C-3 and C-4 have a linear structure, i.e., a single questionnaire results in one record in each file. The C-2 file is different. The C-2 file provides a separate record for each member of a household, so that one questionnaire may result in several C-2 records. The file C-2 could have the same structure as other files but in order to illustrate flexibility of the system, a different solution was pursued.

A record is a set of data about one or several objects. Each record can be subdivided into fields. A field contains a datum detailing some attribute.

Each record, in all four files, bears an identification code that allows the files to be merged into a single file.

Figure 1, below, illustrates that files C-1, C-3, and C-4 have only one record corresponding to a questionnaire while file C-2 contains as many records of a specific questionnaire as there are household members. The figure shows the number of records created from a single questionnaire with two household members. This method of using a separate record for each item guarantees an economical use of the storage area. However, the data stored according to the structure of file C-2 may need to be restructured before it can be used by other software packages.

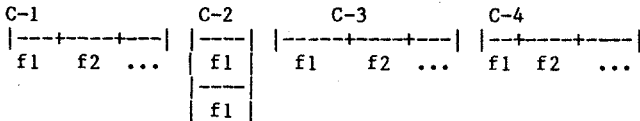


Figure 1. dBASE record structure

f1, f2... depict fields

Figure 2 illustrates the structure of file C-2 following restructuring, which is done by integrating all member-of-the-household records into one record.

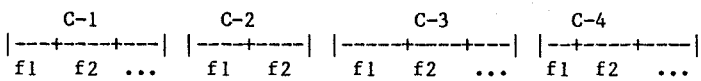


Figure 2. Restructured record for data transfer

The methods used to store data in file C-2 are convenient in a situation with limited data storage and when questions may have a large number of answers with uneven distributions. Thus, because some households may contain two persons and others, 20, the solution presented in Figure 2 would have to reserve space for all 20 possible members, although, on average, a household may have only six members.

3.2.5.1 Combining data files

As noted above, because dBASE III has a maximum of 128 fields per file, the system uses four data files, C-1, C-2, C-3, and C-4.

There are three ways to combine files into one if necessary:

(1) Using dBASE III -

The files can be copied using dBASE III's COPY and UPDATE commands.

(2) Using a high-level language -

A special purpose program written in BASIC, C-language or another high-level language can read and then merge any or all of the data base files.

(3) Using a host package -

The data is transported to a host package (SAS in this case) in dBASE III form. The SAS package thus "imports" dBASE III formatted files, and any additional manipulation of data can be carried out in SAS using the DATA steps.

Methods 1 and 3 are preferred. In both cases the meta information, i.e., variable identification (names) and characteristics, will be passed to the host package.

A file of type C-2 can be aggregated separately, so that from several records only one composite record with totals will be passed to a tabulation stage. One should carefully analyze the tabulation plan in order to pre-aggregate some of the data. If, for example, there is no cross-tabulation on a level of individual household member, then the "TOTAL" command on dBASE III can produce a single record per questionnaire.

3.2.5.2 Creating dBASE III data files with the help of WORDSTAR.

To enter hundreds of variable names may be in itself a cumbersome task. The author uses the following dBASE III command:

```
CREATE <new file> FROM <structure extended file>
```

Where <structure extended file> is a dBASE III file having the predefined structure. <structure extended file> is a type of meta-meta-file common to all files and any file will produce some structure.

This method involves the four steps presented below:

1. Create the structure of a <structure extended file>.

- . USE file (any dBASE file)
- . COPY TO C-STR STRUCTURE EXTENDED.

C-STR.DBF file is created by COPY TO C-STR STRUCTURE EXTENDED.

```
. USE C-STR
. DISP STRU
Structure for database: C:C-STR.dbf
Number of data records:      13
Date of last update   : 08/05/86
Field  Field Name  Type      Width  Dec
  1  FIELD_NAME  Character   10
  2  FIELD_TYPE  Character    1
  3  FIELD_LEN   Numeric     3
  4  FIELD_DEC   Numeric     3
** Total **                18
```

Because the content of this file is irrelevant, the records in C-STR are all deleted by using the ZAP command.

2. Create an ASCII TEXT file with variable specifications. This text file is created using a word processor and contains a list of variables for the target dBASE III file.

```
position
12345678901234567
-----
Q_0_1      N  2  0
Q_0_2      N  2  0
```

```
in position  1-10  variable name
              11   dBASE III data type C,N,D
              12-14 number of characters or total
                   digits, including decimal point
                   (right-aligned)
              15-17 number of decimal digits
```

A list of variables is created using a word processor.

Presented below is the file Y.Y created by WORDSTAR:

```
Q_0_1      N 2 0
Q_0_2      N 2 0
Q_0_3      N 2 0
Q_0_4      N 2 0
Q_3_7_00 AN 2 0
Q_3_7_00 CN 1 0
Q_3_7_00 DN 2 0
Q_3_7_00 EN 1 0
Q_3_7_00 FN 2 0
Q_3_7_00 GN 2 0
Q_3_7_00 IN 1 0
Q_3_7_00 JN 1 0
Q_3_7_00 KN 1 0
```

3. Append variable specifications to the <file C-STR.DBF>.

```
. USE C-STR
. APPEND FROM Y.Y SDF
```

C-STR now contains a list of variables and their specifications

4. Create the database structure.

```
. CREATE C-2 FROM C-STR
```

Finally, file C-2 is created.

```
. DISP STRU
```

Structure for database: C:C-2.dbf

Number of data records: 0

Date of last update : 08/05/86

Field	Field Name	Type	Width	Dec
1	Q_0_1	Numeric	2	
2	Q_0_2	Numeric	2	
3	Q_0_3	Numeric	2	
4	Q_0_4	Numeric	2	
5	Q_3_7_00 A	Numeric	2	
6	Q_3_7_00 C	Numeric	1	
7	Q_3_7_00 D	Numeric	2	
8	Q_3_7_00 E	Numeric	1	
9	Q_3_7_00 F	Numeric	2	
10	Q_3_7_00 G	Numeric	2	
11	Q_3_7_00 I	Numeric	1	
12	Q_3_7_00 J	Numeric	1	
13	Q_3_7_00 K	Numeric	1	
** Total **			22	

The methods described above may be used to modify data structures with a large number of fields. To change a LENGTH-characteristic for 100 variables can be a very time consuming task if the MODIFY STRUCTURE command of dBASE III is used. The method described above is an alternative way to perform the task quickly.

In order to modify in a systematic way a large number of variables the following sequence would be applied:

Assume that characteristics of "afile" are to be modified.

```
USE afile
COPY TO bfile STRUCTURE EXTENDED
USE bfile
COPY TO cfile SDF (&& cfile is a text ASCII file)
QUIT (exit dBASE)
```

```
bring cfile to a wordprocessor
make global changes (replacements)
```

```
dBASE III (reenter dBASE)
USE bfile
ZAP
APPEND FROM cfile SDF
CREATE afile FROM bfile
```

3.3 Editing and correction using dBASE III

Opinions differ on whether editing and correction can be totally incorporated into the data entry stage. The main arguments against data entry combined with editing are:

1. The data entry will be slowed down.
2. Automatic imputation is not easy to implement.

If an extensive error detection programme is implemented, the data should be almost "clean". Elaborate error checking at the data entry stage would probably call for the presence of a subject matter specialist at the data entry site. This specialist would correct errors that call for imputation of data. Alternatively, "bad" entries would be flagged-out and imputations done at the end of the day.

An automatic imputation using the "hot-deck" method may create some problem because a frequency distribution of specific values may be not representative as it is based on a small number of cases.

There are several arguments for incorporating an editing function into the data entry stage:

- (1) Data entry stage is made more efficient -- To search for questionnaires that contain errors at the conclusion of the data entry stage would be time consuming and would require a special physical organization of questionnaires.
- (2) Learning is reinforced -- Data entry staff learn to pay attention to aspects that frequently cause problems.
- (3) Coding is improved -- Systematic errors in coding can be easily discovered at the data entry stage and rectified.

As we have mentioned previously, it is our aim to get "clean" data during the data entry phase. The majority of data editing will be incorporated into the data entry stage.

If, for any reason, some of the errors slipped through the data entry stage, it is possible to use the standard dBASE III commands, LIST or DISPLAY. LIST ALL FOR Q1_4_00Z > 7 will produce a list of records containing values larger than 7, for a variable Q1_4_00Z, i.e., wrong values. Then, the alternations can be done using the dBASE III command EDIT <recno>.

3.4 Tabulation using SAS

3.4.1 Transferring the data from dBASE III to SAS

C-1, C-2, C-3, and C-4 are the data files of the data entry system. To move these data to SAS, the following commands were used:

```
LIBNAME CEN 'NS';
FILENAME DB 'NDBASENC-1.DBF';
PROC DBF DB3=DB OUT=CEN.C1DB;
RUN;
```

```
LIBNAME CEN 'NS';
FILENAME DB 'NDBASENC-2.DBF';
PROC DBF DB3=DB OUT=CEN.C2DB;
RUN;
```

```
LIBNAME CEN 'NS';
FILENAME DB 'NDBASENC-3.DBF';
PROC DBF DB3=DB OUT=CEN.C3DB;
RUN;
```

```
LIBNAME CEN 'NS';
FILENAME DB 'NDBASENC-4.DBF';
PROC DBF DB3=DB OUT=CEN.C4DB;
RUN;
```


After the transfer, the structure of the SAS data set created by moving the C-1 file is as shown below.

Presented here are the first 10 variables of the file C1DB in SAS corresponding to file C-1 in dbase III. The structure of C1DB file is shown in Appendix D. The structure of C-1 is shown in Appendix B; the length of a record is somewhat longer in SAS.

SAS 15:11 Monday, August 11, 1986 1

CONTENTS PROCEDURE

Dsname: CEN.C1DB

Data Set Label:

Number of Observations: 1 Number of Variables: 92

Record Length: 740

Data Set Type:

Memtype: DATA

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Leng	Pos	Format	Informat	Label
1	Q0_1	Num	8	4	10.		
2	Q0_2	Num	8	12	10.		
3	Q0_3	Num	8	20	10.		
4	Q0_4	Num	8	28	10.		
5	Q1_2_00Z	Num	8	36	10.		
6	Q1_4_00Z	Num	8	44	10.		
7	Q1_7_00Z	Num	8	52	10.		
8	Q1_9_00Z	Num	8	60	10.		
9	Q2_1_00Z	Num	8	68	10.		
10	Q2_2_00Z	Num	8	76	10.		

3.4.2 Tabulation variables and their classes

The main tabulation variables include:

- (a) Total holding area (or possibly agricultural land area)
- (b) Cropland (or possibly arable land area) of holding
- (c) Number of cattle on holding
- (d) Number of permanent agricultural workers
- (e) Land tenure
- (f) Holder's legal status
- (g) Size of holder's household
- (h) Holder's age
- (i) Holder's sex
- (j) Intensity of arable land cultivation.

Classes (often called groups) within each variable to be used in the tabulation process are listed in Appendix C.

3.4.3 Generating tables in SAS

At the data processing stage, the data are in SAS dataset format, in one or several files. At this stage it is important that each table be created from variables in a single file. The merging may be general, i.e., one single file is created that contains all the variables and consequently, all the tables can be produced from that file. Alternatively, ad hoc merges can be done which will give input to a set of tables.

Tables can be obtained by writing a SAS program or by filling-in the menu in the full screen option. A sample table produced using the SAS program is presented below.

To obtain a series of similar tables covering different administrative subdivisions of a country, the BY option of SAS can be used.

Survey and sample census processing involves the handling of extrapolation factors. Extrapolation factors can be entered during the data entry phase or added afterwards. They can be stored within each record or added in the tabulation process. It would be safer to insert extrapolation factors after the data entry is done. A simple dBASE III program could do the job automatically, ensuring consistency of the procedure.

The data were moved from dBASE III to SAS using PROC DBF. The transfer routine also moves variable names, thus the table specification is much simplified.

Number and area of holdings by integration with another economic unit
13:06 Friday, August 22, 1986

holding size	Integration with another economic unit					
	engaged in another economic activity					
	yes		no		total	
	number	area (ha)	number	area (ha)	number	area (ha)
without land	19.00	0.00	24.00	0.00	43.00	0.00
0.00- 0.09 ha	23.00	0.59	19.00	0.88	42.00	1.47
0.10- 0.19 ha	15.00	2.15	9.00	1.25	24.00	3.40
0.20- 0.49 ha	76.00	23.80	58.00	19.41	134.00	43.21
0.50- 0.99 ha	30.00	19.02	23.00	17.38	53.00	36.40
1.00- 1.99 ha	28.00	42.94	38.00	45.89	66.00	88.83
2.00- 2.99 ha	90.00	242.04	107.00	268.34	197.00	510.38
3.00- 3.99 ha	55.00	186.40	56.00	186.13	111.00	372.53
4.00- 4.99 ha	6.00	26.12	5.00	21.54	11.00	47.66
5.00- 9.99 ha	53.00	377.70	71.00	491.99	124.00	869.69
10.00- 19.99 ha	10.00	156.76	9.00	153.74	19.00	310.50
20.00- 49.99 ha	60.00	2142.86	90.00	2957.98	150.00	5100.84
50.00- 99.99 ha	3.00	157.62	5.00	275.02	8.00	432.64
100.00- 199.99 ha	.	.	1.00	120.00	1.00	120.00
200.00- 499.99 ha	1.00	250.00	.	.	1.00	250.00
500.00- 999.99 ha	1.00	620.00	.	.	1.00	620.00
1000.00-2499.99 ha	1.00	1100.00	.	.	1.00	1100.00
2500 ha and over	1.00	2750.00	.	.	1.00	2750.00
Total	472.00	8098.00	515.00	4559.55	987.00	12657.55

The program used to generate the sample table is as follows:

```
OPTIONS LS=132 PS=70;
LIBNAME CEN 'NS';
DATA _NULL_ ;
SET CEN.CYDB;
QY = Q4_6_00D;
PROC FORMAT;
  VALUE AREAFMT
    0000.00-0000.00 = 'without land'
    0000.00-0000.09 = ' 0.00- 0.09 ha'
    0000.10-0000.19 = ' 0.10- 0.19 ha'
    0000.20-0000.49 = ' 0.20- 0.49 ha'
    0000.50-0000.99 = ' 0.50- 0.99 ha'
    0001.00-0001.99 = ' 1.00- 1.99 ha'
    0002.00-0002.99 = ' 2.00- 2.99 ha'
    0003.00-0003.99 = ' 3.00- 3.99 ha'
    0004.00-0004.99 = ' 4.00- 4.99 ha'
    0005.00-0009.99 = ' 5.00- 9.99 ha'
    0010.00-0019.99 = ' 10.00- 19.99 ha'
    0020.00-0049.99 = ' 20.00- 49.99 ha'
    0050.00-0099.99 = ' 50.00- 99.99 ha'
    0100.00-0199.99 = ' 100.00- 199.99 ha'
    0200.00-0499.99 = ' 200.00- 499.99 ha'
    0500.00-0999.99 = ' 500.00- 999.99 ha'
    1000.00-2499.99 = '1000.00-2499.99 ha'
    2500.00-9999.99 = '2500 ha and over '
  ;
  VALUE ECONFMT
    1 = 'yes'
    2 = 'no'
  ;
PROC TABULATE;
TITLE 'Number and area of holdings by
      integration with another economic unit';

CLASS Q4_6_00D Q2_2_00Z;
VAR QY;

FORMAT Q4_6_00D AREAFMT.;
FORMAT Q2_2_00Z ECONFMT.;

LABEL
  Q4_6_00D = 'holding size'
  QY       = 'Integration with another economic unit'
  Q2_2_00Z = 'engaged in another economic activity'
  ;
KEYLABEL
  ALL = 'total'
  N   = 'number'
  SUM = 'area (ha)'
  ;
TABLE (Q4_6_00D ALL), QY*(Q2_2_00Z ALL)*(N SUM);
RUN;
```

As noted above, the data entered using the WCA data entry system can be moved to SAS. A new version of SAS 5.0 provides a menu-driven facility to use the TABULATE procedure. After loading SAS, one must press the Ctrl key and A-key simultaneously. The following screen then appears. By entering number 12, which stands for TABULATE, additional screens will appear.

```
+AF=====
| Command ==> 12
|
|           S A S   P R O C E D U R E   M E N U   S Y S T E M
|
|           BASE                Utilities                SAS/STAT
|
|           1 CALENDAR          21 APPEND                41 ANOVA
|           2 CHART             22 COMPARE              42 DISCRIM
|           3 CORR              23 CONTENTS            43 FACTOR
|           4 FORMS            24 COPY                 44 FREQ
|           5 FREQ              25 DATASETS           45 GLM
|           6 MEANS            26 DBF                 46 NPARIWAY
|           7 PLOT             27 DIF                 47 REG
|           8 PRINT            28 FORMAT              48 SCORE
|           9 RANK             29 SORT                 49 TTEST
|          10 STANDARD
|          11 SUMMARY          Remote Link          SAS/IML
|          12 TABULATE         31 DOWNLOAD           51 IML
|          13 TIMEPLOT         32 UPLOAD
|          14 TRANSPOSE        33 SIGNON
|          15 UNIVARIATE       34 SIGNOFF
|=====
```

Menu-screens generated by selecting command no. 12 are found in Appendix E. The fields to be filled-in provide a summary of possible tabulation parameters.

3.5 Analysis and presentation using LOTUS 1-2-3 and SAS

Within SAS the possibilities to change the layout of a table are limited. To make additional aggregations, it may not be worthwhile to use SAS and to run through all data. It thus may be more convenient to move the data to a spreadsheet, i.e., LOTUS 1-2-3. Several methods to accomplish this transfer are detailed below.

3.5.1 Transferring tables from SAS to LOTUS 1-2-3

The table produced by SAS is stored in a temporary OUTPUT file in the ASCII format. To save a file from the OUTPUT WINDOW of SAS one uses the command:

FILE 'destination file name'

The SAS OUTPUT file should be renamed so that it has the extension ".PRN" as LOTUS 1-2-3 expects the file to have this PRN extension.

The file can be imported to LOTUS using this command sequence:

```
/FILE IMPORT (TEXT or NUMBERS)
```

The decision to choose (TEXT or NUMBERS) will depend on the format of the PRN-file.

In the TEXT-mode the table can be directly moved to the spreadsheet but in order to manipulate the data, once in LOTUS, the /DATA PARSE command must be used in order to break ASCII lines into columns.

NUMBERS-mode requires previous manipulation of the table using an editor or wordprocessor. WORDSTAR is a tool for the preparation of the data for submission to LOTUS in NUMBERS-mode. Lines separating data rows were removed, and stubs were surrounded by quotation characters. Titles were also removed.

3.5.2 Manipulation of tables in LOTUS 1-2-3

Using LOTUS, percentages of the totals and a standard deviation were calculated and put in additional columns. The revision of the SAS table is presented on the next page. This is only one example of the additional processing that can be done using LOTUS.

(Pictured below is an actual LOTUS print-out)

Integration with another economic unit
engaged in another economic activity

	yes		no		total		total	
	number	area	number	area	number	area	number	area
without land	19	0	24	0	43	0	4.36%	0.00%
0.00- 0.09 ha	23	0.59	19	0.88	42	1.47	4.26%	0.01%
0.10- 0.19 ha	15	2.15	9	1.25	24	3.4	2.43%	0.03%
0.20- 0.49 ha	76	23.8	58	19.41	134	43.21	13.58%	0.34%
0.50- 0.99 ha	30	19.02	23	17.38	53	36.4	5.37%	0.29%
					296	84.48	29.99%	0.67%
1.00- 1.99 ha	28	42.94	38	45.89	66	88.83	6.69%	0.70%
2.00- 2.99 ha	90	242.04	107	268.34	197	510.38	19.96%	4.03%
3.00- 3.99 ha	55	186.4	56	186.13	111	372.53	11.25%	2.94%
4.00- 4.99 ha	6	26.12	5	21.54	11	47.66	1.11%	0.38%
					385	1019.4	39.01%	8.05%
5.00- 9.99 ha	53	377.7	71	491.99	124	869.69	12.56%	6.87%
10.00- 19.99 ha	10	156.76	9	153.74	19	310.5	1.93%	2.45%
20.00- 49.99 ha	60	2142.86	90	2957.98	150	5100.84	15.20%	40.30%
50.00- 99.99 ha	3	157.62	5	275.02	8	432.64	0.81%	3.42%
					301	6713.67	30.50%	53.04%
100.00- 199.99 ha	0	0	1	120	1	120	0.10%	0.95%
200.00- 499.99 ha	1	250	0	0	1	250	0.10%	1.98%
500.00- 999.99 ha	1	620	0	0	1	620	0.10%	4.90%
1000.00-2499.99 ha	1	1100	0	0	1	1100	0.10%	8.69%
2500 ha and over	1	2750	0	0	1	2750	0.10%	21.73%
					5	4840	0.51%	38.24%
total	472	8098	515	4559.55	987	12657.55	100.00%	100.00%
standard deviation						1730.143		
(derived classes)								
0.00 - 0.99 ha							29.99%	0.6674%
1.00- 4.99 ha							39.01%	8.0537%
5.00- 99.99 ha							30.50%	53.0408%
100 and more ¹							0.51%	38.2380%

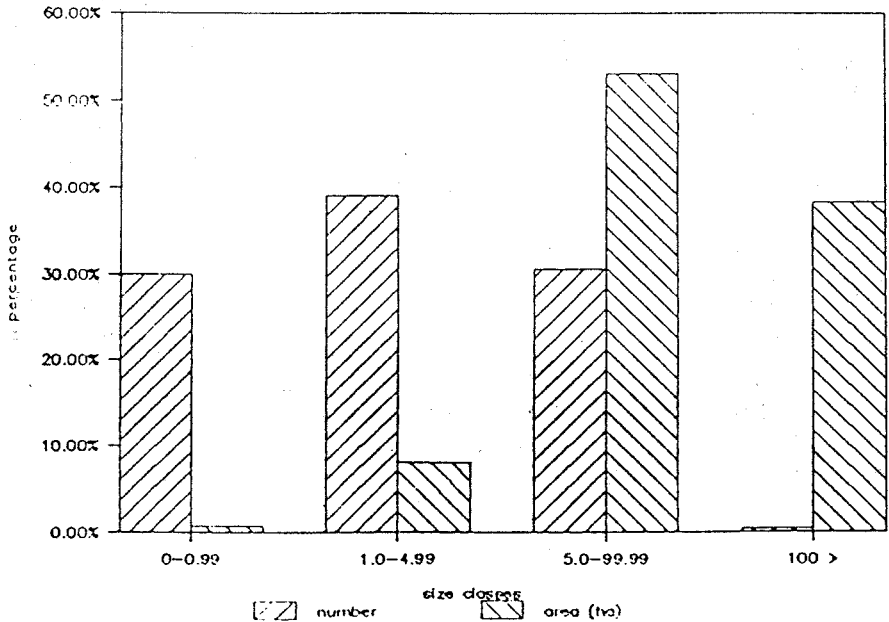
All data included in boxes were calculated using the LOTUS 1-2-3 package on the basis of the original table generated by SAS.

The grouping done on the original data was used to produce the graph which appears on the next page.

3.5.3 Generating graphs in LOTUS 1-2-3

The graph is produced using the LOTUS 1-2-3 graph option and is generated on the basis of the new groupings of the data as specified in LOTUS 1-2-3.

holdings by number and area (in %)



CHAPTER 4

FUTURE DEVELOPMENT

4.1 Software development

The integrated system of dBASE III/dBASE III PLUS, SAS version 6.01 and LOTUS 1-2-3 release 2 reflects the present state of the development of these packages. Later versions of these packages will be released providing even more powerful commands and a more flexible data processing environment. The improvements in the software packages will influence the design of the integrated environment.

dBASE III improvements

The following improvements in dBASE III may influence the design of the data entry system:

- the inclusion of more than 128 fields per file
- a zero response in the RANGE option.

SAS improvements

The latest version of SAS (6.2) is considerably more user-friendly; however, much more should be done to bring SAS to the same level as PC-targeted software.

Multi-user operating system

The major qualitative step will occur when a new multi-user version of MS-DOS is released. Then, several workstations (terminals) can be attached to a single microcomputer. As a bottleneck in the data entry stage is the number of available terminals, the possibility of attaching several workstations (terminals) will have a very positive effect on the total execution of a census or on survey processing.

Compiling dBASE III code

At present, the data entry system needs dBASE III software but there are various possibilities to compile dBASE code to produce executable code. Such a code will reduce per-workstation costs, and the dBASE III itself will not be necessary. However, it is impossible to alter the executable code, so the user must be certain that all of the processing requirements of a data entry system are met before compilation.

4.2 Hardware development

More powerful processors

It can be expected that PCs will continue to develop and "more powerful" models (larger memory, quicker microprocessor) will be introduced. At present, IBM AT uses the 80286 processor; the next version of this processor will be the 80386, which provides the power of a minicomputer.

Larger memory

More primary memory will be put into the microcomputers and the size of secondary storage will be increased.

Networking

DBASE III Plus runs on a network of PCs. The development of Local Area Networks (LANs) will have a major influence on the data entry stage.

How the hardware resources will be utilized depends very much on their development within an operating system. Already available are small and extremely powerful microcomputers that run under UNIX. Unfortunately, at present, there are not enough software packages available under UNIX to cause the larger group of users to switch to UNIX.

(detection of possible errors follows)

```
IF (N 03 = 1 .OR. N 03 = 2) .AND. (N 04 > 0)
  @ 00,00 SAY "1.7 and 1.9 conflicting"
  LOOP
ENDIF
```

(the data is written into a file)

```
REPLACE Q 1 2 00 Z WITH N 01
REPLACE Q 1 4 00 Z WITH N 02
REPLACE Q 1 7 00 Z WITH N 03
REPLACE Q 1 9 00 Z WITH N 04
```

```
M NOK = "0"
ENDDO M NOK
(returning to the main program)
```

RETURN

MAIN MODULE OF THE SYSTEM

*** C-ENTX *****

PROCEDURE C ENT

```
SET HEADING OFF
SET SAFETY OFF
SET TALK OFF
SET BELL OFF
SET STATUS OFF
SET SCOREBOARD OFF
SET TYPEAHEAD TO 0
```

SET CONFIRM ON

*SET ECHO ON

*SET DEBUG ON

*SET BELL OFF

SELECT 1

USE C-1

SELECT 2

USE C-2

SELECT 3

USE C-3

SELECT 4

USE C-4

SELECT 1

CLEAR

X_ANS = " "

M_CONT = "1"

DO WHILE M_CONT = "1"

X_ANS = "N"

@ 00,60 SAY " "

@ 01,00 SAY "Next/Quit/Correct " GET X_ANS PICTURE "!"

READ

IF X_ANS = "Q"

 M_CONT = "0"

 EXIT

ENDIF

IF X_ANS = "C" &&-----correction loop -----

CLEAR

@ 02,00 SAY "=====";
 + "====="

X_CHOICE = " "

@ 06,00 SAY "Select module to "

@ 07,00 SAY "to correct "

@ 08,00 SAY "for example 6.1 " GET X_CHOICE PICTURE "9.9"

READ

IF X_CHOICE > "0.0" .AND. X_CHOICE <= "3.1"

 SELECT 1

ENDIF

IF X_CHOICE = "3.7"

 SELECT 2

ENDIF

IF X_CHOICE >= "3.8" .AND. X_CHOICE <= "5.7"

 SELECT 1

```
*****  
ENDIF  
IF X_CHOICE >= "6.1" .AND. X_CHOICE <= "7.1"  
*****  
SELECT 3  
*****  
ENDIF  
IF X_CHOICE >= "7.2" .AND. X_CHOICE <= "9.1"  
*****  
SELECT 4  
*****  
ENDIF
```

```
DO CASE  
CASE X_CHOICE = "0.1"  
DO CQ01  
CASE X_CHOICE = "1.1"  
DO CQ11  
CASE X_CHOICE = "2.1"  
DO CQ21  
CASE X_CHOICE = "3.1"  
DO CQ31  
CASE X_CHOICE = "3.7"  
DO CQ37  
CASE X_CHOICE = "3.8"  
DO CQ38  
CASE X_CHOICE = "4.1"  
DO CQ41  
CASE X_CHOICE = "4.6"  
DO CQ46  
CASE X_CHOICE = "4.7"  
DO CQ47  
CASE X_CHOICE = "5.1"  
DO CQ51  
CASE X_CHOICE = "5.6"  
DO CQ56  
CASE X_CHOICE = "5.7"  
DO CQ57  
CASE X_CHOICE = "6.1"  
DO CQ61  
CASE X_CHOICE = "7.1"  
DO CQ71  
CASE X_CHOICE = "7.2"  
DO CQ72  
CASE X_CHOICE = "8.1"  
DO CQ81  
CASE X_CHOICE = "9.1"  
DO CQ91  
OTHERWISE
```

ENDCASE

ENDIF && ----- correction loop -----

IF X_ANS = "N" && ----- Next questionnaire -----

SELECT 1

APPEND BLANK

DO CQ01
DO CQ11
DO CQ21
DO CQ31

SELECT 2

DO CQ37

SELECT 1

DO CQ38
DO CQ41
DO CQ46
DO CQ47
DO CQ51
DO CQ55
DO CQ56
DO CQ57

SELECT 3

APPEND BLANK
REPLACE Q_0_1 WITH X_0_1
REPLACE Q_0_2 WITH X_0_2
REPLACE Q_0_3 WITH X_0_3
REPLACE Q_0_4 WITH X_0_4

DO CQ61
DO CQ71

SELECT 4

APPEND BLANK
REPLACE Q_0_1 WITH X_0_1
REPLACE Q_0_2 WITH X_0_2
REPLACE Q_0_3 WITH X_0_3

```
REPLACE Q_0_4      WITH X_0_4
DO CQ72
DO CQ81
DO CQ91
```

```
CLEAR
ENDIF &&----- end next quest-----
```

```
M_CONT = "1"
ENDDO M_CONT
```

```
RETURN
*****
```


APPENDIX B: dBASE III file structures and variables

. USE C-1
. DISP STRU

Structure for database: C:C-1.dbf

Number of data records: 1

Date of last update : 08/11/86

Field	Field Name	Type	Width	Dec
1	Q0_1	Numeric	2	
2	Q0_2	Numeric	2	
3	Q0_3	Numeric	2	
4	Q0_4	Numeric	2	
5	Q1_2_00Z	Numeric	1	
6	Q1_4_00Z	Numeric	1	
7	Q1_7_00Z	Numeric	1	
8	Q1_9_00Z	Numeric	1	
9	Q2_1_00Z	Numeric	1	
10	Q2_2_00Z	Numeric	1	
11	Q2_3_00Z	Numeric	1	
12	Q3_1_00Z	Numeric	2	
13	Q3_2_00Z	Numeric	1	
14	Q3_3_00Z	Numeric	1	
15	Q3_4_00Z	Numeric	1	
16	Q3_5_00Z	Numeric	1	
17	Q3_6_00Z	Numeric	2	
18	Q3_7_00Z	Numeric	1	
19	Q3_8_01Z	Numeric	3	
20	Q3_8_02Z	Numeric	3	
21	Q3_9_00Z	Numeric	1	
22	Q4_1_00Z	Numeric	1	
23	Q4_2_00Z	Numeric	1	
24	Q4_3_00Z	Numeric	1	
25	Q4_4_00Z	Numeric	1	
26	Q4_5_00Z	Numeric	2	
27	Q4_6_00C	Numeric	2	
28	Q4_6_00D	Numeric	6	2
29	Q4_6_00E	Numeric	6	2
30	Q4_6_00F	Numeric	6	2
31	Q4_6_00G	Numeric	6	2
32	Q4_6_00H	Numeric	6	2
33	Q4_6_00I	Numeric	6	2
34	Q4_6_00J	Numeric	6	2
35	Q4_6_00K	Numeric	6	2
36	Q4_6_00L	Numeric	6	2
37	Q4_7_00B	Numeric	6	2
38	Q4_7_00C	Numeric	6	2
39	Q4_7_00D	Numeric	6	2
40	Q4_7_00E	Numeric	6	2
41	Q4_7_00F	Numeric	6	2
42	Q4_7_00G	Numeric	6	2
43	Q4_7_00H	Numeric	6	2

44	Q4 7 00I	Numeric	6	2
45	Q4 7 00J	Numeric	6	2
46	Q4 7 00K	Numeric	6	2
47	Q4 7 00L	Numeric	6	2
48	Q5 1 00Z	Numeric	1	
49	Q5 2 00Z	Numeric	1	
50	Q5 3 00Z	Numeric	1	
51	Q5 4 00Z	Numeric	1	
52	Q5 5 00B	Numeric	6	2
53	Q5 5 00C	Numeric	6	2
54	Q5 5 00D	Numeric	6	2
55	Q5 5 00E	Numeric	6	2
56	Q5 5 00F	Numeric	6	2
57	Q5 5 00G	Numeric	6	2
58	Q5 6 01C	Numeric	2	
59	Q5 6 01D	Numeric	6	2
60	Q5 6 02C	Numeric	2	
61	Q5 6 02D	Numeric	6	2
62	Q5 6 03C	Numeric	2	
63	Q5 6 03D	Numeric	6	2
64	Q5 6 04C	Numeric	2	
65	Q5 6 04D	Numeric	6	2
66	Q5 6 05C	Numeric	2	
67	Q5 6 05D	Numeric	6	2
68	Q5 7 10C	Numeric	2	
69	Q5 7 11F	Numeric	6	2
70	Q5 7 12F	Numeric	6	2
71	Q5 7 13F	Numeric	6	
72	Q5 7 14F	Numeric	6	
73	Q5 7 20C	Numeric	2	
74	Q5 7 21F	Numeric	6	2
75	Q5 7 22F	Numeric	6	2
76	Q5 7 23F	Numeric	6	
77	Q5 7 24F	Numeric	6	
78	Q5 7 30C	Numeric	2	
79	Q5 7 31F	Numeric	6	2
80	Q5 7 32F	Numeric	6	2
81	Q5 7 33F	Numeric	6	
82	Q5 7 34F	Numeric	6	
83	Q5 7 40C	Numeric	2	
84	Q5 7 41F	Numeric	6	2
85	Q5 7 42F	Numeric	6	2
86	Q5 7 43F	Numeric	6	
87	Q5 7 44F	Numeric	6	
88	Q5 7 50C	Numeric	2	
89	Q5 7 51F	Numeric	6	2
90	Q5 7 52F	Numeric	6	2
91	Q5 7 53F	Numeric	6	
92	Q5 7 54F	Numeric	6	
**	Total **		370	

- . USE C-2
- . DISP STRU

Structure for database: C:C-2.dbf

Number of data records: 10

Date of last update : 08/11/86

Field	Field Name	Type	Width	Dec
1	Q0_1	Numeric	2	
2	Q0_2	Numeric	2	
3	Q0_3	Numeric	2	
4	Q0_4	Numeric	2	
5	Q3_7_00A	Numeric	2	
6	Q3_7_00C	Numeric	1	
7	Q3_7_00D	Numeric	2	
8	Q3_7_00E	Numeric	1	
9	Q3_7_00F	Numeric	2	
10	Q3_7_00G	Numeric	2	
11	Q3_7_00I	Numeric	1	
12	Q3_7_00J	Numeric	1	
13	Q3_7_00K	Numeric	1	
** Total **			22	

- . USE C-3
- . DISP STRU

Structure for database: C:C-3.dbf

Number of data records: 1

Date of last update : 08/11/86

Field	Field Name	Type	Width	Dec
1	Q0_1	Numeric	2	
2	Q0_2	Numeric	2	
3	Q0_3	Numeric	2	
4	Q0_4	Numeric	2	
5	Q6_1_00Z	Numeric	1	
6	Q6_2_01C	Numeric	2	
7	Q6_2_01D	Numeric	5	
8	Q6_2_01I	Numeric	1	
9	Q6_2_02C	Numeric	2	
10	Q6_2_02D	Numeric	5	
11	Q6_2_02I	Numeric	1	
12	Q6_2_03C	Numeric	2	
13	Q6_2_03D	Numeric	5	
14	Q6_2_03I	Numeric	1	
15	Q6_2_04C	Numeric	2	
16	Q6_2_04D	Numeric	5	
17	Q6_2_04I	Numeric	1	
18	Q6_2_05C	Numeric	2	
19	Q6_2_05D	Numeric	5	
20	Q6_2_05I	Numeric	1	
21	Q7_1_01C	Numeric	2	
22	Q7_1_01E	Numeric	1	
23	Q7_1_01F	Numeric	2	

24	Q7_1_01G	Numeric	2
25	Q7_1_01H	Numeric	2
26	Q7_1_01I	Numeric	2
27	Q7_1_01J	Numeric	2
28	Q7_1_01K	Numeric	2
29	Q7_1_01L	Numeric	2
30	Q7_1_02C	Numeric	2
31	Q7_1_02E	Numeric	1
32	Q7_1_02F	Numeric	2
33	Q7_1_02G	Numeric	2
34	Q7_1_02H	Numeric	2
35	Q7_1_02I	Numeric	2
36	Q7_1_02J	Numeric	2
37	Q7_1_02K	Numeric	2
38	Q7_1_02L	Numeric	2
39	Q7_1_03C	Numeric	2
40	Q7_1_03E	Numeric	1
41	Q7_1_03F	Numeric	2
42	Q7_1_03G	Numeric	2
43	Q7_1_03H	Numeric	2
44	Q7_1_03I	Numeric	2
45	Q7_1_03J	Numeric	2
46	Q7_1_03K	Numeric	2
47	Q7_1_03L	Numeric	2
48	Q7_1_04C	Numeric	2
49	Q7_1_04E	Numeric	1
50	Q7_1_04F	Numeric	2
51	Q7_1_04G	Numeric	2
52	Q7_1_04H	Numeric	2
53	Q7_1_04I	Numeric	2
54	Q7_1_04J	Numeric	2
55	Q7_1_04K	Numeric	2
56	Q7_1_04L	Numeric	2
57	Q7_1_05C	Numeric	2
58	Q7_1_05E	Numeric	1
59	Q7_1_05F	Numeric	2
60	Q7_1_05G	Numeric	2
61	Q7_1_05H	Numeric	2
62	Q7_1_05I	Numeric	2
63	Q7_1_05J	Numeric	2
64	Q7_1_05K	Numeric	2
65	Q7_1_05L	Numeric	2
66	Q7_1_06C	Numeric	2
67	Q7_1_06E	Numeric	1
68	Q7_1_06F	Numeric	2
69	Q7_1_06G	Numeric	2
70	Q7_1_06H	Numeric	2
71	Q7_1_06I	Numeric	2
72	Q7_1_06J	Numeric	2
73	Q7_1_06K	Numeric	2

74	Q7_1_06L	Numeric	2
75	Q7_1_07C	Numeric	2
76	Q7_1_07E	Numeric	1
77	Q7_1_07F	Numeric	2
78	Q7_1_07G	Numeric	2
79	Q7_1_07H	Numeric	2
80	Q7_1_07I	Numeric	2
81	Q7_1_07J	Numeric	2
82	Q7_1_07K	Numeric	2
83	Q7_1_07L	Numeric	2
** Total **			169

. USE C-4

. DISP STRU

Structure for database: C:C-4.dbf

Number of data records: 1

Date of last update : 08/11/86

Field	Field Name	Type	Width	Dec
1	Q0_1	Numeric	2	
2	Q0_2	Numeric	2	
3	Q0_3	Numeric	2	
4	Q0_4	Numeric	2	
5	Q7_2_01C	Numeric	2	
6	Q7_2_01E	Numeric	1	
7	Q7_2_01F	Numeric	2	
8	Q7_2_01G	Numeric	2	
9	Q7_2_01H	Numeric	2	
10	Q7_2_01I	Numeric	2	
11	Q7_2_01J	Numeric	2	
12	Q7_2_01K	Numeric	2	
13	Q7_2_01L	Numeric	2	
14	Q7_2_02C	Numeric	2	
15	Q7_2_02E	Numeric	1	
16	Q7_2_02F	Numeric	2	
17	Q7_2_02G	Numeric	2	
18	Q7_2_02H	Numeric	2	
19	Q7_2_02I	Numeric	2	
20	Q7_2_02J	Numeric	2	
21	Q7_2_02K	Numeric	2	
22	Q7_2_02L	Numeric	2	
23	Q7_2_03C	Numeric	2	
24	Q7_2_03E	Numeric	1	
25	Q7_2_03F	Numeric	2	
26	Q7_2_03G	Numeric	2	
27	Q7_2_03H	Numeric	2	
28	Q7_2_03I	Numeric	2	
29	Q7_2_03J	Numeric	2	
30	Q7_2_03K	Numeric	2	
31	Q7_2_03L	Numeric	2	
32	Q7_2_04C	Numeric	2	
33	Q7_2_04E	Numeric	1	
34	Q7_2_04F	Numeric	2	
35	Q7_2_04G	Numeric	2	
36	Q7_2_04H	Numeric	2	
37	Q7_2_04I	Numeric	2	
38	Q7_2_04J	Numeric	2	

39	Q7 2 04K	Numeric	2
40	Q7 2 04L	Numeric	2
41	Q7 2 05C	Numeric	2
42	Q7 2 05E	Numeric	1
43	Q7 2 05F	Numeric	2
44	Q7 2 05G	Numeric	2
45	Q7 2 05H	Numeric	2
46	Q7 2 05I	Numeric	2
47	Q7 2 05J	Numeric	2
48	Q7 2 05K	Numeric	2
49	Q7 2 05L	Numeric	2
50	Q7 2 06C	Numeric	2
51	Q7 2 06E	Numeric	1
52	Q7 2 06F	Numeric	2
53	Q7 2 06G	Numeric	2
54	Q7 2 06H	Numeric	2
55	Q7 2 06I	Numeric	2
56	Q7 2 06J	Numeric	2
57	Q7 2 06K	Numeric	2
58	Q7 2 06L	Numeric	2
59	Q7 2 07C	Numeric	2
60	Q7 2 07E	Numeric	1
61	Q7 2 07F	Numeric	2
62	Q7 2 07G	Numeric	2
63	Q7 2 07H	Numeric	2
64	Q7 2 07I	Numeric	2
65	Q7 2 07J	Numeric	2
66	Q7 2 07K	Numeric	2
67	Q7 2 07L	Numeric	2
68	Q8 1 00Z	Numeric	1
69	Q8 2 01G	Numeric	1
70	Q8 2 01H	Numeric	4
71	Q8 2 01I	Numeric	5
72	Q8 2 02G	Numeric	1
73	Q8 2 02H	Numeric	4
74	Q8 2 02I	Numeric	5
75	Q8 2 03G	Numeric	1
76	Q8 2 03H	Numeric	4
77	Q8 2 03I	Numeric	5
78	Q8 2 04G	Numeric	1
79	Q8 2 04H	Numeric	4
80	Q8 2 04I	Numeric	5
81	Q8 2 05G	Numeric	1
82	Q8 2 05H	Numeric	4
83	Q8 2 05I	Numeric	5
84	Q8 2 06G	Numeric	1
85	Q8 2 06H	Numeric	4
86	Q8 2 06I	Numeric	5
87	Q9 1 06Z	Numeric	1
88	Q9 2 06Z	Numeric	5
89	Q9 3 06Z	Numeric	1
90	Q9 4 06Z	Numeric	1
**	Total	**	197

APPENDIX C: Class descriptions for the main variables

Classes for main variables

Total area of holding (051102)

This classification covers all holdings. The same classification is recommended for agricultural land area.

- (a) Holdings without land
- (b) Holdings with land
 - Under 0.1 ha
 - 0.1 ha and under 0.2 ha
 - 0.2 ha and under 0.5 ha
 - 0.5 ha and under 1 ha

 - 1 ha and under 2 ha
 - 2 ha and under 3 ha
 - 3 ha and under 4 ha
 - 4 ha and under 5 ha
 - 5 ha and under 10 ha

 - 10 ha and under 20 ha
 - 20 ha and under 50 ha
 - 50 ha and under 100 ha

 - 100 ha and under 200 ha
 - 200 ha and under 500 ha
 - 500 ha and under 1000 ha
 - 1000 ha and under 2500 ha
 - 2500 ha and over

Cropland of holding (05241, 05242)

This classification covers all holdings. The same classification is recommended for arable land area.

- (a) Holdings without cropland
- (b) Holdings with cropland
 - Under 0.1 ha
 - 0.1 ha and under 0.2 ha
 - 0.2 ha and under 0.5 ha
 - 0.5 ha and under 1 ha

 - 1 ha and under 2 ha
 - 2 ha and under 3 ha
 - 3 ha and under 4 ha
 - 4 ha and under 5 ha
 - 5 ha and under 10 ha

10 ha and under 20 ha
20 ha and under 50 ha
50 ha and under 100 ha

100 ha and under 200 ha
200 ha and under 500 ha
500 ha and under 1000 ha
1000 ha and over

Number of cattle of holding (071201)

This classification covers all holdings

- (a) Holdings with no cattle
- (b) Holdings with cattle

- 1 to 2 head
 - 3 to 4 head
 - 5 to 9 head

- 10 to 19 head
 - 20 to 49 head
 - 50 to 99 head

- 100 to 199 head
 - 200 to 499 head
 - 500 head and over

Number of permanent agricultural workers (042221,
042222)

This classification covers all holdings

- (a) Holdings without permanent agricultural workers
- (b) Holdings with permanent agricultural workers

- 1 worker
 - 2 workers
 - 3 workers
 - 4 workers
 - 5 workers
 - 6 workers and over

Land tenure (0522)

This classification covers all holdings. The break down (c) may not be applicable in some cross tabulations.

- (a) Holdings without land

- (b) Holdings operated under one tenure form
Owned or held in ownerlike possession
Rented from others
- For an agreed amount of money and/or produce
 - For a share of produce
 - In exchange for services
 - Under other rental arrangements
- Operated on a squatter basis
Operated under tribal or traditional communal tenure forms
Operated under other tenure forms
- (c) Holdings operated under two or more tenure forms
Area owned or held in ownerlike possession
Area rented from others
- For an agreed amount of money and/or produce
 - For a share of produce
 - In exchange for services
 - Under other rental arrangements
- Area operated on a squatter basis
Area operated under tribal or traditional communal tenure forms
Area operated under other tenure forms

Legal status of holder (021101)

This classification covers all holdings.

- (a) Holdings operated privately
- An individual
 - A household
 - Two or more individuals of different households or two or more households
 - A corporation
 - A cooperative
 - Other n.e.s
- (b) Holdings operated by the Government

Holder's household size (031101)

This classification covers only those holdings operated by an individual or a household. Other holdings are excluded.

- 1 person
- 2 to 3 persons
- 4 to 5 persons
- 6 to 9 persons
- 10 persons and over

Holder's age (031202)

This classification applies only to holdings operated by an individual, a household, two or more individuals of different households or two or more households. It excludes other holdings. In the case of holdings operated by a household, two or more individuals of different households or two or more households, one senior holder should be selected.

Under 25 years
25 to 34 years
35 to 44 years
45 to 54 years
55 to 64 years
65 years and over

Holder's sex (031203)

Explanations given above for holder's age, equally apply here.

Male
Female

Arable land cultivation intensity (05241, 061141)
This classification applies only to those holdings having arable land.

Under 0.50
0.50 and under 0.75
0.75 and under 1.00
1.00 and under 1.25
1.25 and under 1.50
1.50 and under 2.00
2.00 and over

Classes for other characteristics

Category 01

Administrative units and agro-ecological regions (011101)

Classes by administrative units and agro-ecological zones are to be determined by countries.

Category 02

Integration with another economic unit (021301)
This classification covers all holdings.

Holdings that are part of an enterprise also engaged in another economic activity.

Holdings that are not part of an enterprise also engaged in another economic activity.

Other economic activities of enterprises operating a holding (021302)

This classification applies to holdings that are part of an enterprise also engaged in another economic activity.

- (a) Single economic activity other than operating a holding
 - Agricultural services
 - Hunting, trapping and game propagation
 - Forestry and logging
 - Fishing
 - Manufacturing
 - Other (specify)

- (b) Two or more economic activities other than operating a holding

Hired manager (021201, 021204)

Holdings with a hired manager

Holdings without a hired manager

Category 03

Age of holder's household members (031202)

The first three age groups are combined for the holder when cross-classifying holder's age (see Holder's age).

- Under 10 years
- 10 to 14 years
- 15 to 24 years
- 25 to 34 years
- 35 to 44 years
- 45 to 54 years
- 55 to 64 years
- 65 years and over

Sex of holder's household members (0310203)

Male

Female

Category 04

Activity status (041101)

This classification covers holder's household members above a specified age - see paragraph 3.41.

Economically active

Not economically active

Number of holder's household members economically active (041101)

- 1 person
- 2 persons
- 3 persons
- 4 persons
- 5 persons or more

Work of household members on holding (042101, 042102)

This classification applies to economically active members of holder's household.

- Occasional agricultural worker on holding
- Permanent agricultural worker on holding
- Not worked on holding

Holder's occupations (041201, 041202)

- Not having any occupation besides being a holder
- Having other occupation(s) besides being a holder, with main occupation of
 - Farmer
 - Agricultural worker
 - Other (specify)

Employment of agricultural workers (042211, 042212)

- Holdings that employ agricultural workers only on a permanent basis
- Holdings that employ agricultural workers only on an occasional basis
- Holdings that employ agricultural workers both on permanent and occasional basis
- Holdings that do not employ agricultural workers
- Sex of permanent agricultural workers (042221, 042222)

- Male
- Female

Category 05

Number of parcels (051101)

This classification applies to holdings with land.

- 1 parcel
- 2 to 3 parcels
- 4 to 5 parcels
- 6 to 9 parcels
- 10 parcels and more

Land rented out (051211)

- Holdings that rent out land
- Holdings that do not rent out land

Shifting cultivation (051401)

- Holdings practising shifting cultivation
- Holdings not practising shifting cultivation

Irrigation (051812)

- Holdings that irrigate some land
- Holdings that do not irrigate any land

Drainage (051911)

- Holdings having drainage facilities
- Holdings not having drainage facilities

Size of parcels (052101)

- Under 0.1 ha
- 0.1 ha and under 0.2 ha
- 0.2 ha and under 0.5 ha
- 0.5 ha and under 1 ha
- 1 ha and under 2 ha
- 2 ha and under 5 ha
- 5 ha and under 10 ha
- 10 ha and over

Land use (0521, 0524)

(a) Agricultural land

Cropland

Arable land

- Area under temporary crops - in the open air
- Area under temporary meadows
- Area temporarily fallow
- All other arable land

Land under protective cover

Land under permanent crops - in the open air

Land under permanent meadows and pastures

Cultivated

Naturally grown

(b) Woodland or forest

(c) All other land

- Unused and undeveloped potentially productive land,
- Land in the holding n.e.s.

(d) Total land

Time elapsed since clearance for cultivation (in shifting cultivation) (0521, 0526)

- Area cleared less than 1 year ago
- Area cleared 1 year and less than 2 years ago
- Area cleared 2 years and less than 5 years ago
- Area cleared 5 years ago or earlier

Category 06

Area harvested (061141)
For each temporary crop grown
Under 1 ha
1 ha and under 2 ha
2 ha and under 5 ha
5 ha and under 10 ha
10 ha and under 20 ha
20 ha and over

Plantation features and ages of permanent crops (061241-061245)

For each permanent crop
Trees of productive age in compact plantations
Trees of non-productive age in compact plantations
Scattered trees

Fertilizers (0613)

- (a) Holdings that do not use fertilizers
- (b) Holdings that use fertilizers
 - Only inorganic fertilizers
 - Only organic fertilizers
 - Both types of fertilizers

Pesticides (0614)

- Holdings that do not use pesticides
- Holdings that use pesticides

Seeds (0615)

- Holdings that use high yield variety seeds
- Holdings that do not use high yield variety seeds

Crop stand (0521, 0621)

- Area under pure stand - temporary crops
- Area under pure stand - permanent crops
- Area under mixed - temporary crops
- Area under mixed - permanent crops
- Area under associated crops

Category 07-08

Livestock system (071101)

- Nomadic or totally pastoral
- Semi-nomadic or semi-pastoral
- Sedentary pastoral
- Ranching

Number of buffaloes (071202)

1 to 2 head
3 to 4 head
5 to 9 head
10 to 19 head
20 to 49 head
50 to 99 head
100 to 199 head
200 to 499 head
500 head and over

Number of sheep (071203)

1 to 4 head
5 to 9 head
10 to 19 head
20 to 49 head
50 to 99 head
100 to 199 head
200 to 499 head
500 head and over

Number of goats (071204)

1 to 4 head
5 to 9 head
10 to 19 head
20 to 49 head
50 to 99 head
100 to 199 head
200 to 499 head
500 head and over

Number of pigs (071205)

1 to 4 head
5 to 9 head
10 to 19 head
20 to 49 head
50 to 99 head
100 to 199 head
200 to 499 head
500 head and over

Number of chickens (071301)

1 to 9 chickens
10 to 49 chickens
50 to 199 chickens
200 to 999 chickens
1000 to 4999 chickens
5000 to 9999 chickens
10 000 chicken and over

Number of deer (071303)

- 1 to 4 head
- 5 to 9 head
- 10 to 19 head
- 20 to 49 head
- 50 to 99 head
- 100 to 199 head
- 200 to 499 head
- 500 head and over

Category 09

Indication of machinery and equipment used and source for each type (091101, 091201)

- (a) Holdings that do not use machinery and equipment
- (b) Holdings that do use machinery and equipment
 - Owned solely by holder
 - Owned jointly by holder and others
 - Provided by landlord
 - Provided by other private holders (excluding cooperatives)
 - Provided by private cooperative
 - Provided by a private agricultural service establishment
 - Provided by a government agency

Category 10

Nonresidential buildings use (102111)

- Holdings that use nonresidential buildings
- Holdings that do not use nonresidential buildings

Use of nonresidential buildings owned (102121, 10221)

- For keeping livestock, other than poultry
- For keeping poultry
- For storing agricultural products
- For keeping agricultural machinery and equipment
- As office
- For mixed or other purposes

Category 11

Existence and number of forest trees (111101, 111211)

- Holdings with no forest trees
- Holdings with forest trees

- 1 to 9 trees
- 10 to 49 trees
- 50 to 199 trees
- 200 trees and over

Fisheries activity and cultural installation (112101, 112201)

- Holdings that do not carry out fisheries activity
- Holdings that carry out fisheries activity
 - in pond
 - in rice fields
 - in other cultural installations

APPENDIX D: File structure of a sample file in SAS

SAS 15:11 Monday, August 11, 1986 1

CONTENTS PROCEDURE

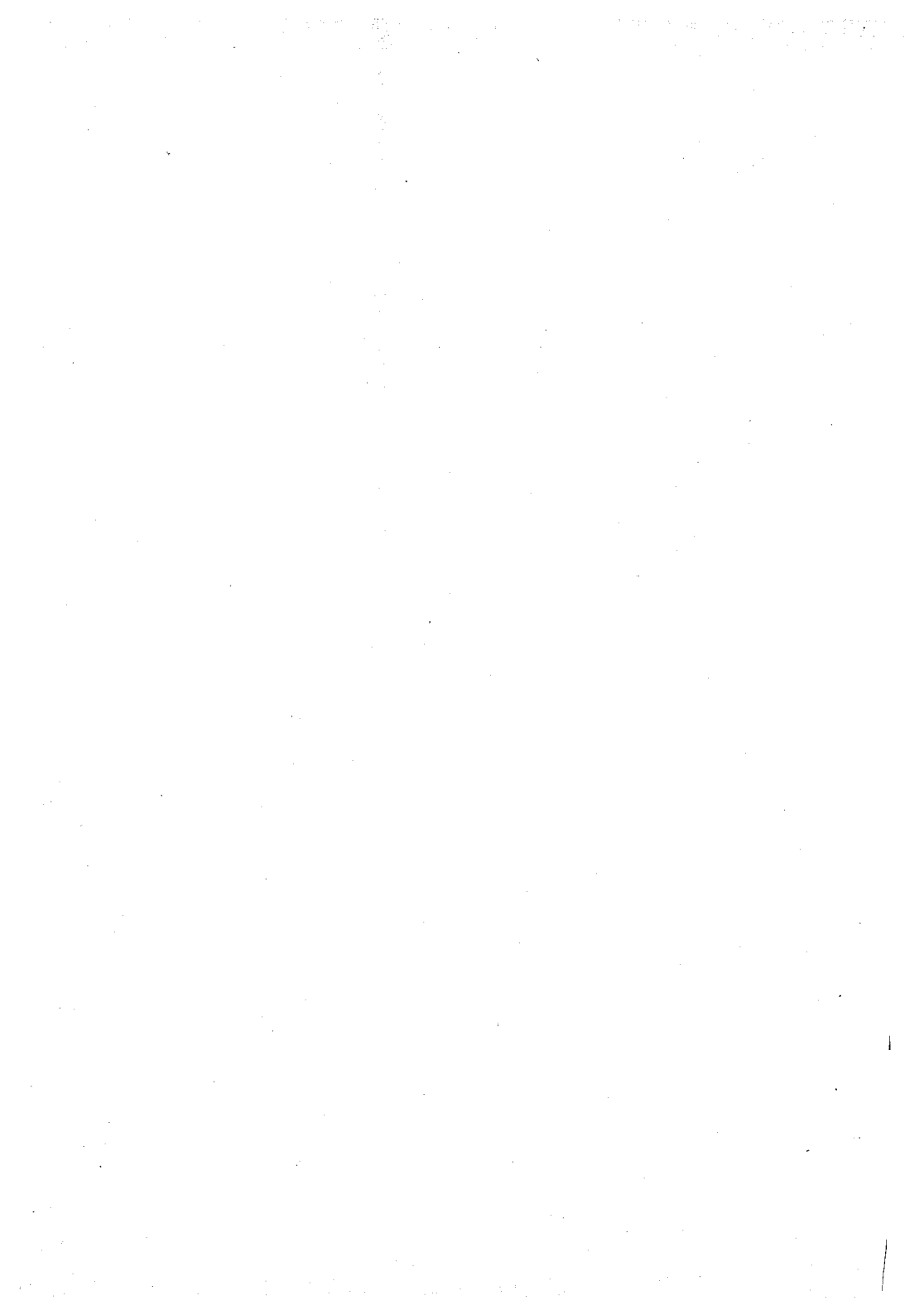
Dsname: CEN.C1DB
Data Set Label:
Number of Observations: 1 Number of Variables: 92
Record Length: 740
Data Set Type:
Mentype: DATA

-----Alphabetic List of Variables and Attributes-----

#	Variable	Type	Leng	Pos	Format	Informat	Label
1	Q0_1	Num	8	4	10.		
2	Q0_2	Num	8	12	10.		
3	Q0_3	Num	8	20	10.		
4	Q0_4	Num	8	28	10.		
5	Q1_2_00Z	Num	8	36	10.		
6	Q1_4_00Z	Num	8	44	10.		
7	Q1_7_00Z	Num	8	52	10.		
8	Q1_9_00Z	Num	8	60	10.		
9	Q2_1_00Z	Num	8	68	10.		
10	Q2_2_00Z	Num	8	76	10.		
11	Q2_3_00Z	Num	8	84	10.		
12	Q3_1_00Z	Num	8	92	10.		
13	Q3_2_00Z	Num	8	100	10.		
14	Q3_3_00Z	Num	8	108	10.		
15	Q3_4_00Z	Num	8	116	10.		
16	Q3_5_00Z	Num	8	124	10.		
17	Q3_6_00Z	Num	8	132	10.		
18	Q3_7_00Z	Num	8	140	10.		
19	Q3_8_01Z	Num	8	148	10.		
20	Q3_8_02Z	Num	8	156	10.		
21	Q3_9_00Z	Num	8	164	10.		
22	Q4_1_00Z	Num	8	172	10.		
23	Q4_2_00Z	Num	8	180	10.		
24	Q4_3_00Z	Num	8	188	10.		
25	Q4_4_00Z	Num	8	196	10.		
26	Q4_5_00Z	Num	8	204	10.		
27	Q4_6_00C	Num	8	212	10.		
28	Q4_6_00D	Num	8	220	10.2		
29	Q4_6_00E	Num	8	228	10.2		
30	Q4_6_00F	Num	8	236	10.2		

31	Q4_6_00G	Num	8	244	10.2
32	Q4_6_00H	Num	8	252	10.2
33	Q4_6_00I	Num	8	260	10.2
34	Q4_6_00J	Num	8	268	10.2
35	Q4_6_00K	Num	8	276	10.2
36	Q4_6_00L	Num	8	284	10.2
37	Q4_7_00B	Num	8	292	10.2
38	Q4_7_00C	Num	8	300	10.2
39	Q4_7_00D	Num	8	308	10.2
40	Q4_7_00E	Num	8	316	10.2
41	Q4_7_00F	Num	8	324	10.2
42	Q4_7_00G	Num	8	332	10.2
43	Q4_7_00H	Num	8	340	10.2
44	Q4_7_00I	Num	8	348	10.2
45	Q4_7_00J	Num	8	356	10.2
46	Q4_7_00K	Num	8	364	10.2
47	Q4_7_00L	Num	8	372	10.2
48	Q5_1_00Z	Num	8	380	10.
49	Q5_2_00Z	Num	8	388	10.
50	Q5_3_00Z	Num	8	396	10.
51	Q5_4_00Z	Num	8	404	10.
52	Q5_5_00B	Num	8	412	10.2
53	Q5_5_00C	Num	8	420	10.2
54	Q5_5_00D	Num	8	428	10.2
55	Q5_5_00E	Num	8	436	10.2
56	Q5_5_00F	Num	8	444	10.2
57	Q5_5_00G	Num	8	452	10.2
58	Q5_6_01C	Num	8	460	10.
59	Q5_6_01D	Num	8	468	10.2
60	Q5_6_02C	Num	8	476	10.
61	Q5_6_02D	Num	8	484	10.2
62	Q5_6_03C	Num	8	492	10.
63	Q5_6_03D	Num	8	500	10.2
64	Q5_6_04C	Num	8	508	10.
65	Q5_6_04D	Num	8	516	10.2
66	Q5_6_05C	Num	8	524	10.
67	Q5_6_05D	Num	8	532	10.2
68	Q5_7_10C	Num	8	540	10.
69	Q5_7_11F	Num	8	548	10.2
70	Q5_7_12F	Num	8	556	10.2
71	Q5_7_13F	Num	8	564	10.
72	Q5_7_14F	Num	8	572	10.
73	Q5_7_20C	Num	8	580	10.
74	Q5_7_21F	Num	8	588	10.2
75	Q5_7_22F	Num	8	596	10.2
76	Q5_7_23F	Num	8	604	10.
77	Q5_7_24F	Num	8	612	10.
78	Q5_7_30C	Num	8	620	10.
79	Q5_7_31F	Num	8	628	10.2
80	Q5_7_32F	Num	8	636	10.2
81	Q5_7_33F	Num	8	644	10.
82	Q5_7_34F	Num	8	652	10.
83	Q5_7_40C	Num	8	660	10.

84	Q5_7_41F	Num	8	668	10.2
85	Q5_7_42F	Num	8	676	10.2
86	Q5_7_43F	Num	8	684	10.
87	Q5_7_44F	Num	8	692	10.
88	Q5_7_50C	Num	8	700	10.
89	Q5_7_51F	Num	8	708	10.2
90	Q5_7_52F	Num	8	716	10.2
91	Q5_7_53F	Num	8	724	10.
92	Q5_7_54F	Num	8	732	10.



APPENDIX E: Panels for PROC TABULATE from the SAS AF Window.

```
AF
Command ==>
Use Scroll commands for more.

PROC TABULATE DATA= _____
  _ MISSING _ NOSEPS FORMAT= _____ DEPTH= ____
  ORDER= INTERNAL or _ FREQ or _ DATA or _ FORMATTED
  FORMCHAR( _____ ) = ' _____ ' ;

CLASS ..... ;
VAR ..... ;
FREQ ..... ;
WEIGHT ..... ;
BY ..... ;
```

```
AF
Command ==>
Use Scroll commands for more.

TABLE (page) (row) (column) / options;
Expressions may include MEAN, SUM, VAR, N, SUMWGT, MAX, MIN.

TABLE Page expression _____ ,
      Row expression _____ ,
      Column expression _____ /

TABLE statement options:
  _ PRINTMISS _ CONDENSE MISSTEXT= ' _____ '
ROW= CONSTANT or _ FLOAT RTSPACE= ____
FUZZ= _____
BOX= _ _PAGE_ or BOX= _____ or BOX= ' _____ ' ;
```

```
AF-
Command ==>

                                Use Scroll commands for more.
STATEMENTS TO LABEL VARIABLES AND SPECIFY FORMATS

FORMAT   variable           format.
FORMAT   .....            -----
FORMAT   .....            -----
FORMAT   .....            -----
FORMAT   .....            -----

LABEL    variable           = label
LABEL    .....            = '
          .....            = '
          .....            = '
          .....            = '

KEYLABEL keyword= 'text'
          -----
          -----
          -----
          -----
```

C

APPENDIX F:

REPUBLIC OF AGRARIA

Central Statistical Office

QUESTIONNAIRE

CENSUS OF AGRICULTURE

1990

CONFIDENTIALITY: The data provided by the respondent in this questionnaire are strictly confidential. They will be used only for statistical purposes.

OBLIGATION : All persons (civil or juridical) living in the country are obliged by law to provide the data requested in this questionnaire.

OBJECTIVE : The objective of the census of agriculture is to collect basic data on agricultural structure to be used in preparing policies for the development of agriculture.

IDENTIFICATION OF THE QUESTIONNAIRE

	Code
0.1 Province	<input type="text"/>
0.2 District	<input type="text"/>
0.3 Enumeration area	<input type="text"/>
0.4 Serial number of the questionnaire	<input type="text"/>
0.5 Expansion factor	<input type="text"/>

Name of enumerator
Date of enumeration

Name of supervisor
Date of control

Name of coder
Date coded

PART 1 IDENTIFICATION OF HOLDER, RESPONDENT, HIRED MANAGER

A. Holder, respondent

1.1 Name of the holder

.....

1.2 Is the holder living in the locality where the holding is?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

(If yes, pass to question 1.4; if no, continue with question 1.3.)

1.3 Address of the holder

.....

1.4 The respondent, is he/she the holder?

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

(If yes, pass to question 1.7; if no, continue with question 1.5.)

1.5 Name of the respondent

.....

1.6 Relationship of the respondent to the holder

.....

B. Hired Manager

1.7 Who runs the daily work of the holding?
(Put x mark in appropriate box.)

Holder him (her) self	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Joint holder	<input type="checkbox"/>	<input checked="" type="checkbox"/>
Hired manager	<input type="checkbox"/>	<input checked="" type="checkbox"/>

(If 1 or 2, pass to question 2.1; if 3, continue with question 1.8.)

1.8 Name of hired manager

.....

1.9 Does the hired manager live on the holding?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

(If yes, pass to question 2.1; if no, continue with question 1.10.)

1.10 Address of the hired manager

.....

PART 2 GENERAL CHARACTERISTICS

2.1 Is the holder

an individual
a household,
two or more individuals of different
households or two or more households,
a corporation,
a cooperative,
another private organization,
the Government?
(Put x mark in appropriate box.)

<input type="checkbox"/>	1
<input type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input type="checkbox"/>	6
<input type="checkbox"/>	7

2.2 Is the holding part of an enterprise engaged also in other economic activities?
(Put x mark in appropriate box.)

<input type="checkbox"/>	Yes	<input type="checkbox"/>	1
<input type="checkbox"/>	No	<input type="checkbox"/>	2

(If yes, continue with question 2.3; if no, pass to question 3.1.)

2.3 What are the other economic activities of the enterprise?
[Put x mark(s) in appropriate box(es).]

Agricultural services
Hunting, trapping and game propagation
Forestry and logging
Fishing
Manufacturing
Other (specify)

<input type="checkbox"/>	1
<input type="checkbox"/>	2
<input type="checkbox"/>	3
<input type="checkbox"/>	4
<input type="checkbox"/>	5
<input type="checkbox"/>	6

PART 3 DEMOGRAPHIC CHARACTERISTICS AND EMPLOYMENT

A. Holder and household

(Ask these questions only to civil person holders. In case of a household holder ask them to the head of the household. In case of joint holders ask them to all joint holders and fill in a separate PART 3-A for each one)

3.1 Age of the holder

<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------

3.2 Sex of the holder

(Put x mark in appropriate box)

<input type="checkbox"/>	Male	<input type="checkbox"/>	1
<input type="checkbox"/>	Female	<input type="checkbox"/>	2

3.3 Did you have other occupations during the agricultural year, besides being a holder?

(Put x mark in appropriate box.)

<input type="checkbox"/>	Yes	<input type="checkbox"/>	1
<input type="checkbox"/>	No	<input type="checkbox"/>	2

(If yes, continue with question 3.4; if no, pass to question 3.5.)

3.4 Which occupation was your main occupation?
(Put x mark in appropriate box. In case 3, also specify)

Holder	<input type="checkbox"/>	1
Agricultural worker	<input type="checkbox"/>	2
Other	<input type="checkbox"/>	3

3.5 Who is the head of your household?
(Put x mark in appropriate box. In case 2, also indicate name.)

Holder	<input type="checkbox"/>	1	Name.....
Other	<input type="checkbox"/>	2	

3.6 What is the total number of members of the household including the holder?

<input type="checkbox"/>	<input type="checkbox"/>
--------------------------	--------------------------

3.7 Indicate for each member of household, excluding the holder

Serial No.	Name	Sex (put x mark)		Age	For those above 11 years						Codes					
					Whether worked? (put x mark)		If has worked			If has not worked			Eco-Act Status (1) active (2) non-active	Work on holding (1) Perm. worker (2) Occasional (3) Has not worked		
					Yes	No	No. of months			Looked for a job? (put x mark)					Yes	No
							on holding	outside holding	Total	Yes	No					
		M	F		Yes	No	on holding	outside holding	Total	Yes	No					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					
		<input type="checkbox"/>	<input type="checkbox"/>		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>					

B. Hired workers
(Ask these questions to all holders)

3.8 Did you employ any permanent agricultural workers during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	1
No	<input type="checkbox"/>	2

(If yes, continue with question 3.9; if no, pass to question 3.10.)

3.9 Number of permanent agricultural workers
1- Male
2- Female
3- Total

3.10 Did you employ any occasional agricultural workers during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	1
No	<input type="checkbox"/>	2

PART 4 LAND AND WATER

4.1 Do you practice shifting cultivation?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	1
No	<input type="checkbox"/>	2

4.2 Have you rented out any of your land to others during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	1
No	<input type="checkbox"/>	2

4.3 Have you actually irrigated any land in your holding during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	1
No	<input type="checkbox"/>	2

4.4 Are there any drainage facilities for any land in your holding?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	1
No	<input type="checkbox"/>	2

4.5 How many parcels are there in your holding?

<input type="text"/>

PART 5 CROPS

5.1 Have you used any high yield variety seeds during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

5.2 Have you used any inorganic fertilizers during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

5.3 Have you used any organic manure during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

5.4 Have you used any pesticides during the agricultural year?
(Put x mark in appropriate box.)

Yes	<input type="checkbox"/>	<input checked="" type="checkbox"/>
No	<input type="checkbox"/>	<input checked="" type="checkbox"/>

5.5 Indicate the area under various crop stand categories for each parcel.
(Express areas in hectares with two decimals.)

No. of parcel	Total area	Crop stand (ha)				
		Pure stand		Mixed crops		Associated crops
		Temporary crops	Permanent crops	Temporary crops	Permanent crops	
a	b	c	d	e	f	g
1						
2						
3						
4						
5						
6						
7						
8						
9						
10						
Total						

B. Agricultural machinery

7.2 Indicate for each type of agricultural machinery whether used during the agricultural year, and if used, its number by source.

Serial No.	Agricultural Machinery	Code	Used or not (put x mark in appropriate box)	Code	Source (indicate number of machinery)						
					Owned solely by holder	Owned jointly by holder and others	Provided by Landlord	Provided by other private holders (excl. cooperatives)	Provided by cooperatives	Provided by a private agricultural service establishment	Provided by a government agency
a	b	c	d	e	f	g	h	i	j	k	l
1			Yes	1							
			No	2							
2			Yes	1							
			No	2							
3			Yes	1							
			No	2							
4			Yes	1							
			No	2							
5			Yes	1							
			No	2							
6			Yes	1							
			No	2							
7			Yes	1							
			No	2							
8			Yes	1							
			No	2							
9			Yes	1							
			No	2							
10			Yes	1							
			No	2							
11			Yes	1							
			No	2							
12			Yes	1							
			No	2							

PART 8 NONRESIDENTIAL BUILDINGS

8.1 Have you used any nonresidential building for agricultural purposes during the agricultural year?
(Put x mark in appropriate box.)

Yes	1
No	2

(If yes, continue with question 8.2; if no, pass to question 9.1.)

8.2 Indicate, for each nonresidential building in each type class used, its tenure; and for each building owned, its area (or volume).

Serial No. a	Type Class b	Serial No. c	Tenure (*)			For buildings owned		
			Owned d	Rented e	Other f	Code g	Area (m ²) (for other purposes)	Volume (m ³) (for storage)
							h	i
1	For keeping livestock other than poultry	1						
		2						
		3						
		4						
		Total						
2	For keeping poultry	1						
		2						
		3						
		Total						
3	For storing agricultural products	1						
		2						
		3						
		4						
		Total						
4	For keeping agricultural machinery	1						
		2						
		3						
		4						
		Total						
5	As office	1						
		2						
		3						
		Total						
6	For mixed or other purposes	1						
		2						
		3						
		4						
		5						
		Total						

*Put x mark for individual buildings and the number for "Total".

PART 9 ANCILLARY ACTIVITIES

A. Forestry

9.1 Are there any forest trees on the holding?
(Put x mark in appropriate box.)

Yes	1
No	2

(If yes, continue with question 9.2; if no, pass to question 9.3.)

9.2 Total number of forest trees on the holding

--	--	--	--	--	--

B. Fisheries

9.3 Have you taken any fish or other aquatic animals and plants from the waters within the holding during the agricultural year?
(Put x mark in appropriate box.)

Yes	1
No	2

(If yes, continue with question 9.4, if no, terminate interview.)

9.4 Indicate the type of place where fish or other aquatic animals and plants are taken.

[Put x mark(s) in appropriate box(es).]

Pond	1
Rice field	2
Other	3

