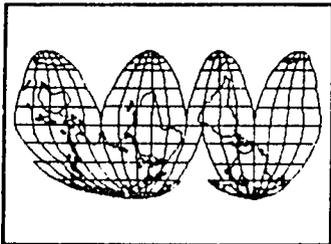


PN-ABS-706



J.E. AUSTIN ASSOCIATES, INC.
1816 Jefferson Place, NW
Washington, DC 20036 USA
tel 202-833-4902 fax 202-833-4904

SPSS MANUAL:

CAMEROON PRIVATE SECTOR DATABASE

Prepared for:

USAID/CAMEROON
Yaounde, Cameroon

October, 1991

I. Introduction to the MAPS database	1
A. Background: why the data was collected	1
B. How the information was collected	1
C. How accurate is this database?	3
D. What kind of information does the database contain?	4
II. Making Use of the MAPS Database	6
A. Welcome to SPSS/PC+	6
B. COMMANDS: We are in charge, not SPSS!	7
C. Getting Started: Retrieving the Cameroon database.	8
III. Practice Exercises: Analyzing Data	13
A. Frequencies	13
B. Cross Tabulations	14
C. Combining data into new groups	16
D. Testing ideas using SPSS	22
E. Three-way Contingency: thinness of data	25
IV. Graphics	30
A. Plotting Frequencies	30
SCATTERPLOT SYMBOLS FOR MULTIPLE CASES	33
B. FASTGRAF	34
1. Setting up the computer and printer	34
2. Using FASTGRAF.	34
V. REGRESSIONS	35
VI. Troubleshooting	41
THE COMPUTER SAYS IT CAN'T DO WHAT I WANT	41
NOTHING HAPPENS!	41
THE WHOLE THING TAKES A LONG TIME!	41
THE GRAPHICS THING WON'T WORK!	41
VII. How to <u>Import</u> the database into other software	42
A. Preparing your data	42
list of the .SYS files and the variables each contains	42
B. Exporting	42
C. Importing	43
VIII. LOTUS 123, and EXCEL FOR WINDOWS	43
IX. Definitions: a glossary	44
APPENDIX A: CAMEROON SURVEY QUESTIONNAIRE AND FREQUENCY DISTRIBUTIONS	49
APPENDIX B - List of Variables used in the Cameroon Survey	76

I. Introduction to the MAPS database

A. Background: why the data was collected

USAID/Cameroon's programs seek to help attain sustainable and broadly based economic growth. One strategic objective is to assist in the development of an efficient and competitive private sector. To do so, the Mission requires valid data upon which to base its decisions. J.E. Austin Associates (JAA) was asked to survey 408 enterprises in Cameroon and provide that data for use in designing and monitoring its private sector assistance program.

The objectives of this survey were:

1. to develop a base of descriptive statistics on private enterprises operating in Cameroon, and,
2. to develop information on the perceptions of Cameroon's business community regarding:
 - a. the past, present and future investment climate
 - b. resource constraints to business development
 - c. the effects of public policies on business growth
 - d. the role and effectiveness of business associations
 - e. interest in export and other investments

The survey yielded a rich database on private enterprises in Cameroon. That database, and how it can help you, is the focus of our attention. As you will read, the data provide a reliable source of primary information on many aspects of company operations and entrepreneurs' aspirations. We have confidence in the reliability of the survey results and in the ability of analysts to extrapolate from the sample to the universe of firms operating in Cameroon.

B. How the information was collected

Working with Input Output Computer Services (IOCS) in Douala, JAA designed a survey instrument to gather data from all major sectors in Cameroon's economy. This survey was not a census of all enterprises, but rather a carefully designed sample intended to reflect the reality of the Cameroon business world at the time the survey was undertaken, February 1990.

1. The Sampling Frame

A key to gathering reliable information is the sampling technique. Before firms can be chosen through such techniques, a database of all firms must be constructed. This is the "universe" or the "sampling frame." Separate sampling frames¹ were constructed for formal and informal sector firms.

a. Formal Sector. IOCS constructed a comprehensive sampling frame based on the most recent census of industry and the DSF (fiscal and statistical declaration forms). The Chamber of Commerce directory and the yellow pages provided telephone numbers and addresses for all the firms drawn for the sample.

b. Informal Sector. Approximately 25 percent of the total sample was reserved for non registered or informal sector firms. Because informal sector firms by definition do not appear in any formal registry, a sampling frame for them is more difficult to design. The survey of informal sector firms was confined to Douala and Yaounde in order to exercise closer quality control on both the sampling and the interviewing process for these firms. The proportion of informal firms drawn from each city was equal approximately to the ratio of formal sector firms in Douala to those in Yaounde, or 60:40. Thus it was assumed that the distribution of informal sector firms in these cities was similar to that of formal sector firms.

The sample of informal firms was drawn using a geographic cluster sampling technique. IOCS field supervisors first established and made a list of the geographical areas, or clusters, in each city where informal sector firms operate. A sample of clusters was drawn from that list. Each cluster was then broken down into blocks. Supervisors assigned to each enumerator the firms to interview within each block. They assigned those firms in such a way that the enumerators chose a sample within each cluster roughly representative of the sectoral distribution of firms in that cluster. This reduced enumerator discretion in choosing the firms, and the likelihood of systematic biases introduced by the manner in which they chose the firms to interview.

¹The sampling frame excluded all firms in which the Government of Cameroon had majority ownership. In cases where this could not be determined a priori, a question in the survey asked firms to specify what percentage of shares (where applicable) were owned by GRC. On average GRC owned 23 percent of the shares of the enterprises included in the sample (refer to Appendix A, question 24). Moreover, the sampling focussed on only major urban areas (See Appendix A, Question 2 for a list of the cities where the survey was conducted).

Since the sampling was concentrated in urban centers, and excluded all subsistence agriculture, the final sample of firms chosen understates the importance of agriculture in Cameroon's economy. Also, in the final distribution of firms by sector of economic activity an attempt was made to include the main industrial sectors in the sample in the same proportion found in each sampling frame. This principle was not applied, however, when the number of firms called for in the sample was less than 5 units. In such cases, at least 6 firms from that industry sector were included. A unit was then subtracted from the other industrial sectors at a correct statistical proportion. Hence, the sectoral distribution in the sample deviates somewhat from the sectoral distribution in the sampling frames, and probably understates the importance of commerce sector firms in the private sector.

C. How accurate is this database?

The sampling procedures described are designed to minimize bias in the responses. For formal sector firms, the random sampling process used means we can be fairly confident that the responses accurately reflect the universe of formal sector firms in the country and can even estimate our level of confidence. But, because the sample of the informal sector firms was a nonprobability sample, we cannot generalize the results to the universe of informal sector firms in Cameroon. We have no mathematical theory to estimate the chance of any unit in the universe of being selected into the sample, and, therefore, no basis for estimating population parameters. On the other hand, by minimizing biases when choosing the respondents, we can be fairly confident that the responses were a "good" (though not "statistically valid") representation of the population in the clusters chosen.

Moreover, the rejection rate for this survey exercise was fairly low (less than 15 percent). Of those, roughly one third of the enterprises originally chosen could not be interviewed because they had been closed down. The remainder could not be located because, though their addresses were P.O. Boxes in Douala and Yaounde, these firms were actually located far outside these cities, or because they did not appear in either the Yellow Pages of the phone book or in the Chamber of Commerce registry. But, in every case, IOCS sought to replace firms which could not be interviewed with firms of similar characteristics. The composition of the sample, both sectorally and in terms of size and location, ultimately remained unchanged. This gives the analyst greater confidence in the reliability of the survey results and in the ability to extrapolate from the sample to the universe of firms operating in the country.

D. What kind of information does the database contain?

This database contains information expressed as 364 variables. Each variable corresponds directly to questions about the private sector in Cameroon contained in the survey questionnaire (attached as Appendix A). In addition, there are 17 variables created as composites of two or more of the questions. A list and description of all variables is attached as Appendix B.

There are two basic types of variables: categorical and continuous. A categorical variable is one in which the possible responses are categories. From the actual questionnaire, here is an example of a categorical variable:

5. What is the gender of the owners (or the majority shareholders) of this firm (CIRCLE ONE RESPONSE)
- (1). Male (2). Female (3). Equal proportion of (4). Don't know
shares held by male
and female

Typical categorical variable responses are Male/Female, Yes/No, Very Important/Somewhat Important/Not Important. You will see that, when the responses are coded for the database, these categories will receive numerical values, so that Male will equal 1.00 and Female 2.00.

A continuous variable represents actual figures, such as we find in Question 19:

19. What percentage of your raw materials (approximately) come from:
- | | |
|-------------------------------------------|--------|
| a). Within the province | _____% |
| b). Outside province, but within Cameroon | _____% |
| c). within Other UDEAC ¹ | _____% |
| d). other parts of Africa | _____% |
| e). EEC | _____% |
| f). Other Europe | _____% |
| g). North America | _____% |
| h). Asia | _____% |
| i). Middle East | _____% |
| j). Other | _____% |

Responses to this question will be a numerical value somewhere between 0 and 100. Other examples of a continuous

¹Includes Cameroon, Chad, CAR, Equatorial Guinea, Gabon, Congo.

variable are the total number of employees, or the total number of female employees.

As you have gathered, variables can be simple, and correspond directly to a particular question, or, composite variables, which are the results of a combination of simple variables. Referring to Appendix B, 88., the new variable

V90 (PERCENTAGE WOMEN UNSKILLED):

is a composite variable, created as follows:

COMPUTE V90=V37c3/V37c1

in other words, this new variable represents the value of variable V37c3 (Women, skilled labor) divided by variable V37c1 (total unskilled), to yield the percentage of women, unskilled.

The survey (see Appendix A) contains questions on the general business climate, perceived resource constraints in the areas of credit, labor, energy, transportation, communications and land. It also contained questions on the effectiveness of the support provided by existing business associations. It contains information on where and from whom Cameroonian firms purchase their inputs and to whom they sell their production.

The exercises provided in this manual are designed to illustrate how you can use this database to obtain useful answers to questions you may have about Cameroon's private sector. Examples could include: from which countries do Cameroonian textiles manufacturers buy their inputs? What is the level of demand for which types of business services? What is the nature of present dissatisfaction with business services being offered?

Now turn to Appendix A and spend some time reading through the questions so you will be familiar with the nature of the information the database contains. This will give you insight into how you can manipulate the data to answer your particular questions about the nature of the private sector in Cameroon.

II. Making Use of the MAPS Database

This training manual assumes each user has a basic knowledge of English, statistics and how computer software is, in general, operated.

The responses to the survey questions were entered into a computer database with software widely used by statisticians and analysts since the days of punch cards: SPSS. That software has been updated for the personal computer in a version called SPSS/PC+, which is one of the software tools we will be using to extract useful information from the mass of numbers and responses. (Lotus 1,2,3 and Excel are other tools, about which more later)

A. Welcome to SPSS/PC+

SPSS/PC+ is a system that accesses and analyzes data using simple, consistent language which is easy to learn and does not require much knowledge about computers. SPSS/PC+ writes six types of files:

- the scratch pad - this contains the commands you type or paste into it. Its file name is SCRATCH.PAD; it is saved automatically whenever you leave REVIEW, and when you exit SPSS/PC+
- listing files - these contain your display output. By default, this output is sent to the screen and to the SPSS.LIS file. You can specify a special name for this listing file and also a different disc drive in which to keep it. This can be important, as the SPSS.LIS file is overwritten with new information each time we start a new session, and the old listing will disappear forever.
- log files - this contains a log or record of the commands executed by SPSS/PC+, as well as messages about the commands that caused errors, and about the location of command output in this listing file. The default log is called SPSS.LOG, but, again, we can specify a special name for the log each time and also a different disc drive in which to keep it, for the same reasons as noted above.
- system files - these save data and a data dictionary for use in subsequent SPSS/PC+ sessions, if you use the SAVE command. The default system file, when you use the SAVE command without specifying another filename, is SPSS.SYS. Note that we will be using a series of *.SYS files in this database, named OUTPUT1.SYS through OUTPUT7.SYS. These contain the Cameroon database.

- portable files - are used to transport data and a data dictionary across machines, when you use the EXPORT command, as described below in Chapter VI.

- and, results from procedures - the results files include, naturally, results, such as matrix materials, new data, and so on, produced from specific commands, such as WRITE, CORRELATION, CLUSTER, REGRESSION, and so on. The default results file is called SPSS.PRC.

At the beginning of each SPSS session, the default files SCRATCH.PAD, SPSS.LIS, and SPSS.LOG are reinitialized. In plain language, that means that whatever had been stored in those files will be erased and replaced by new information. If you want to keep the contents of those files, first copy them to a new disc and rename them!

B. COMMANDS: We are in charge, not SPSS!

We will communicate with SPSS/PC+ with simple, English-like statements, known as COMMANDS.

A command can be simple:

LIST.

It can contain additional specifications:

LIST CASES FROM 1 TO 100 BY 5.

Or, a command can contain complex combinations of several subcommands such as this one:

FREQUENCIES VARIABLES = SEX JOBCAT /BARChart /STATISTICS ALL.

This line told the software to generate the frequencies for two variables, Sex and job category, and then to display the results as a barchart labeled with the numbers. Notice and remember that in SPSS/PC+, a command line always ends with a period(.). Without a period, the software will be confused and do nothing.

COMMANDS can be used to do three things, usually in this order:

1. CONVERT DATA into a form that can be used by the computer and bring the data into SPSS/PC+. The data may exist on paper, or in some other electronic form such as DBase or Lotus. Data definition commands, like DATA LIST, TRANSLATE, and GET (among others) tell SPSS/PC+ where and how to read data.

2. MODIFY DATA. Once the data are in SPSS/PC+, we might wish to convert revenues expressed in one currency to revenues measured in US\$. For this we would use a data manipulation command like

COMPUTE REVENUE = REVENUE/270

This command assumes that REVENUE is a variable containing information expressed in CFA and that the exchange rate is CFA270 = US\$1. This command told the software to convert the data by dividing ("/") each REVENUE entry by 270. Naturally, if the exchange rate were different, we would have divided by a different number.

3. PROCESS DATA. When the data are in the form we want, procedure commands like FREQUENCIES, MEANS, and REGRESSION tell SPSS/PC+ to do something with the data, such as perform a statistical analysis, produce a report, a listing, a plot, sort the cases into a different order, or save the data to a FILE.

4. During the course of an SPSS/PC+ session, we can repeat any of these three steps, to get different data, arrange the data differently, or run different analyses on the data.

C. Getting Started: Retrieving the Cameroon database.

1. The System Files

Data are contained in "system" files in SPSS. These files have the suffix .SYS. There were so many variables in this exercise that the compilers created seven different *.SYS files to contain them. These seven *.SYS files were given the names OUTPUT1.SYS through OUTPUT7.SYS. Although we can analyze data in each of these files separately, SPSS will not permit us to crosstabulate the data from one *.SYS file with that of another. To do this we first have to join all the *.SYS files into one large file. Therefore, before we begin to learn how to analyze the data, we will wish to join all seven *.SYS files into one big SYS file in order to extract all possible combinations of data. Because of limitations of memory space, we will do this in several steps, as outlined below.

a. First we must enter the SPSS/PC+ program. Therefore, at the cursor command line C> enter these letters: **SPSSPC** and press **ENTER**.

b. After some machine noises, this will produce the SPSS/PC+ logo screen and then two screens:

MAIN MENU (on the left) and orientation (on the right).

c. Using the down arrow cursor control, place the highlighted bar over line labeled "session control & info", and press ENTER.

d. Move the cursor down to the word SET and press ENTER. You will see the word SET appear on the lower part of the screen.

e. Now you will see the choice 'output'; put the cursor over that word and press ENTER.

f. A new selection will appear. Move the cursor down to /LISTING and press ENTER. Now the lower part of the screen will hold: SET /LISTING

g. Press ENTER again and a new box will appear. Type A:\CAM1.LIS and press ENTER. What you have typed will appear on the lower part of the screen between single quotation marks:

```
SET /LISTING 'A:\CAM1.LIS'.
```

h. Press F10 and a message appears at the very bottom of the screen: Run: Run from Cursor. Press ENTER in order to tell SPSS to carry out this command line. Every action in SPSS/PC+ is activated by the combination of F10 plus ENTER.

We just told SPSS that the file CAM1.LIS, located on a separate disc in the A drive, will be the place to save a complete history, or a listing, of what we are doing: commands and resulting data. In case we make a mistake, this record will be useful to figure out what went wrong and how to correct the mistake. It is also a handy place to store data output to recover later.

i. Now we can begin assembling the seven OUTPUT.SYS files into a single package by typing the following commands. Remember that each line must end with a period (.) for SPSS to act on the command. Still in the 'session control & info' section, press ENTER, move the cursor to GET and press ENTER. Now Press Alt-T (to enable you just to type your command line) and type the following: /FILE 'OUTPUT1.SYS'. and press ENTER. Notice the period (.) at the end of the line. You have to type that, too. This command line should appear on

the lower part of the screen, just below the SET /LISTING 'A:\CAM1.LIS'. command line.

GET /FILE 'OUTPUT1.SYS'. [again touch F10 and ENTER]

A command like SET or GET is always followed by a space; the thing that it is to "set" or "get" is always preceded by a "/". In this case we are telling SPSS to put the first of the seven *.SYS files which contain the data, into its operating memory. Now press ALT-T again and type in the next line:

**JOIN MATCH /FILE * /FILE 'OUTPUT2.SYS' /FILE 'OUTPUT3.SYS'.
[F10 plus ENTER]**

Now we are splicing three *.SYS files together. We tried this at home and discovered that our computer can handle no more than three of these *.SYS files at once. The MATCH command tells SPSS to look for markers at the start of each *.SYS file as points of alignment so that the data will actually correspond to the right questions.

SAVE /OUTFILE 'NEWCAM1.SYS'. [F10 plus ENTER]

Now we ordered SPSS/PC+ to SAVE the results of its work and to call the brand new *.SYS file we created out of the first three of the seven "NEWCAM1.SYS". SPSS knows this as an OUT[put]FILE. Your memory disc now ought to contain NEWCAM1.SYS, the length of which should be the sum of the lengths of OUTPUT1.SYS plus OUTPUT2.SYS plus OUTPUT3.SYS.

This same procedure was performed on the OUTPUT4.SYS, OUTPUT5.SYS and OUTPUT6.SYS files to create another file called NEWCAM2.SYS. Finally, NEWCAM1.SYS and NEWCAM2.SYS can be joined with the remaining OUTPUT7.SYS file. After this is done, you will have all seven of the original files combined into one single SYS file. This will enable you to access all 364 variables at one time. To save time and effort, we have provided as part of this manual a permanent CAMEROON.LOG file that automatically performs all of these steps for you. By running CAMEROON.LOG. you will load SPSS's operating memory with the entire database.

Here is what that program (called CAMEROON.LOG) looks like:

```
GET /FILE 'OUTPUT1.SYS'.  
JOIN MATCH /FILE * /FILE 'OUTPUT2.SYS' /FILE 'OUTPUT3.SYS' /FILE  
SAVE /OUTFILE 'NEWCAM1.SYS'.  
GET /FILE 'OUTPUT4.SYS'.  
JOIN MATCH /FILE * /FILE 'OUTPUT5.SYS' /FILE 'OUTPUT6.SYS'.  
SAVE /OUTFILE 'NEWCAM2.SYS'.  
GET /FILE 'NEWCAM1.SYS'.  
JOIN MATCH /FILE * /FILE 'NEWCAM2.SYS' /FILE 'OUTPUT7.SYS'.
```

To make this work the easy way each time, all you have to do is to issue SPSS the command, `INCLUDE /CAMEROON.LOG.`, from the 'session control & info' line. We have put this program into your computer for you. Notice that the command line ends with a period (.).

Now press the ESC key until you see the selection 'session control & info' appear on your screen. Place the cursor over that line, and issue this command:

```
INCLUDE /CAMEROON.LOG. [press F10 and ENTER]
```

The activity will carry on the screen for a while. Whenever you hear a BEEP and see MORE in the upper right hand corner of the screen, just press the SPACE BAR to tell the computer to carry on.

At this point, assuming you have carried out this procedure, your memory will contain all seven OUTPUT.SYS files as well as NEWCAM1.SYS and NEWCAM2.SYS: a very large chunk of memory.

After the session, remember to copy all seven OUTPUT.SYS files into a separate floppy disc and put it safely away. With NEWCAM1.SYS and NEWCAM2.SYS in your memory, you can delete OUTPUT1.SYS through OUTPUT6.SYS (leaving OUTPUT7.SYS along with NEWCAM1.SYS and NEWCAM2.SYS). That will clear out precious memory, and permit us to write a shorter version of CAMEROON.LOG. to use later, ie:

```
GET /FILE 'NEWCAM1.SYS'.  
JOIN MATCH /FILE * /FILE 'NEWCAM2.SYS' /FILE 'OUTPUT7.SYS'.
```

At this point, SPSS/PC+ is prepared to start extracting data, performing statistical analysis, or make charts for you.

To make everything perfectly clear, here is a listing of an SPSS/PC+ session performed in September, 1991:

INCLUDE 'cameroon.log'.
GET /FILE 'NEWCAM1.SYS'.
The SPSS/PC+ system file is read from
file NEWCAM1.SYS
The file was created on 9/23/91 at 15:46:13
and is titled SPSS/PC+
The SPSS/PC+ system file contains
408 cases, each consisting of
106 variables (including system variables).
106 variables will be used in this session.

-----Page 2
SPSS/PC+ 9/24/91

This procedure was completed at 17:26:55
JOIN MATCH /FILE * /FILE 'NEWCAM2.SYS' /FILE 'OUTPUT7.SYS'.

-----Page 3
SPSS/PC+ 9/24/91

This procedure was completed at 17:30:59

III. Practice Exercises: Analyzing Data

Presuming that you have actually followed each step in the previous chapter, all variables in the Cameroon private sector database are now loaded into SPSS/PC's active memory and available to be analyzed. In the next sections, we will invoke the SPSS/PC+ commands which will order a number of statistical calculations. Again, we are presuming a basic knowledge of statistics.

A. Frequencies

Just to warm up, let's do a simple frequency table of Variable #2. We will start out by pressing the ESC key until we see the main SPSS menu, which contains the choice 'analyze data'. Place the cursor over that line and press ENTER. Now you have new choices. Place the cursor over 'descriptive statistics' and press ENTER. Now you have new choices. Place the cursor over FREQUENCIES and press ENTER. Place the cursor over the choice /VARIABLES and press ENTER. Now the menu tells you to "pop up" the complete choice of variables by pressing Alt-V. Do it. A new box should appear: using the arrow keys, move the cursor over V2 and press ENTER.

Now the command line on the lower part of the screen should show this:

```
FREQUENCIES /VARIABLES V2 /FORMAT DFREQ. [press F10 and ENTER]
```

This command line orders the software to get, compute and print the number of responses, and the percentages, to VARIABLE 2, V2, or Question 2 on the survey: where is your enterprise located. The next illustration shows what your screen should display, eventually.

***** Memory allows a total of 9934 Values, accumulated across all Variables.
 There also may be up to 1241 Value Labels for each Variable.

 Page 5 SPSS/PC+ 10/25/91
 V2 Location

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
Douala	1	221	54.2	54.2	54.2
Yaounde	5	110	27.0	27.0	81.1
Bamenda	8	21	5.1	5.1	86.3
Bafoussam	6	20	4.9	4.9	91.2
Garoua	7	14	3.4	3.4	94.6
Ngaoundere	9	12	2.9	2.9	97.5
Limbe	2	7	1.7	1.7	99.3
Tiko	3	2	.5	.5	99.8
Edea	4	1	.2	.2	100.0

 Valid Cases 408 TOTAL 408 100.0 100.0
 Missing Cases 0

[Therefore we know that 56.4% of all respondents were located in Douala, Limbe, Tiko and Edea. We also know that 27% were located in Yaounde.]

B. Cross Tabulations

Now let's see how many male owned and female owned enterprises were located in each city. We do that with the command which follows See if you can construct it by yourself. If you get confused, study the examples SPSS/PC+ offers on the screen.

CROSSTABS /TABLES V2 BY V5 /OPTIONS 5.

[remember to touch F10 and ENTER in order to activate the FINISH. command]

***** Given WORKSPACE allows for 7285 Cells with
2 Dimensions for CROSSTAB problem *****

Page 20 SPSS/PC+ 10/25/91
Crosstabulation: V2 Location
By V5 Gender-owner

Page 1 of 3

V5→	Count Tot Pct	Male 1	Female 2	Equal pr oportio n 3	Don"t kn ow 4	Row Total
V2						
Douala	1 185 45.3		14 3.4	9 2.2	13 3.2	221 54.2
Limbe	2 4 1.0			2 .5	1 .2	7 1.7
Tiko	3 2 .5					2 .5
Edea	4 1 .2					1 .2
Column (Continued) Total		345 84.6	28 6.9	18 4.4	17 4.2	408 100.0

Crosstabulation: V2 Location
By V5 Gender-owner

Page 2 of 3

V5→	Count Tot Pct	Male 1	Female 2	Equal pr oportio n 3	Don"t kn ow 4	Row Total
V2						
Yaounde	5 95 23.3		11 2.7	3 .7	1 .2	110 27.0
Bafoussam	6 20 4.9					20 4.9
Garoua	7 12 2.9			2 .5		14 3.4
Bamenda	8 15 3.7		2 .5	2 .5	2 .5	21 5.1
Column (Continued) Total		345 84.6	28 6.9	18 4.4	17 4.2	408 100.0

Crosstabulation: V2 Location
By V5 Gender-owner

Page 3 of 3

V5→	Count Tot Pct	Male 1	Female 2	Equal pr oportio n 3	Don"t kn ow 4	Row Total
V2						
Ngaoundere	9 11 2.7		1 .2			12 2.9
Column Total		345 84.6	28 6.9	18 4.4	17 4.2	408 100.0

Number of Missing Observations = 0

That was a sample of a short and simple session with SPSS/PC+. Now look at the Survey questions and the VARIABLES they correspond to and see how you would like to manipulate the data to get an answer to your question.

C. Combining data into new groups

What if the selection of variables is so broad, and each contains such a wide range of data, that a crosstabulation becomes unwieldy? With the relatively simple crosstabulation above, we still generated three separate frames of data.

For example, suppose you want to know which sectors of industry are purchasing no inputs from sources in their province, which are buying 1-50% of inputs from the province and which are buying a majority of inputs from sources located in their province. For suppliers, this information will give an indication of opportunity, while others may interpret the results in terms of efficiencies, costs, constraints, etc.

So let's give SPSS/PC+ some instructions to create a few categories out of a wide range of values contained in a continuous variable. We will select Question 19A: what percentage of raw materials do you buy from sources within your province? This is variable V19A.

The first category may be all those who said they buy a majority of their raw materials from sources within the province. First we select from the main menu the line 'modify data or files' and press ENTER.

Now select the new choice 'modify data' and press ENTER. There are a number of choices with which to start.

Place the cursor over IF and press ENTER. Now press Alt-T and type in the box that pops up: (V19A GT 50) PROV = 3. and press ENTER.

The command line on the lower part of the screen should show this:

```
IF (V19A GT 50) PROV = 3.
```

This command line manipulates data buried in the files. It tells SPSS that IF the value assigned to VARIABLE V19A (what percentage of your inputs do you buy within your Province) is Greater Than 50 (GT) then it will create a new variable, which we have called PROV, with the value of 3. The same holds for the other cases:

```
IF (V19A EQ 0) PROV = 1.
```

[if the value equals 0, ie, no purchases from within the Province, then the new variable PROV will have a value of 1.]

```
IF (V19A GT 0 and V19A LE50) PROV = 2.
```

[note first we have created a categorical variable from a continuous variable]

Now press Alt-E and using the arrow keys move the cursor to the first command line you typed. Press Alt-E again, press F10 and press ENTER.

```
ERROR      222, TEXT: LE50
UNDEFINED VARIABLE NAME ON TRANSFORMATION--Check for a
misspelled variable name.
This command not executed.
```

[oops! a typing error: we forgot to put a space between LE and 50, and SPSS didn't know what to do. Reenter the command properly:]

```
IF (V19A GT 0 and V19A LE 50) PROV = 2.
```

Now press Alt-E, move the cursor up to the first of these three lines again, press Alt-E again, press F10 and press ENTER.

This will order SPSS to modify the values of continuous variable V19A into three categories in variable PROV.

Now that the new variable PROV has been created, we can use it like any of the original variables. Let's begin with a simple frequency analysis, which should reveal the number of responses in each of the three categories, and the percentage of all responses each represents.

Look back at the earlier section on FREQUENCIES and see if you can do this operation by yourself.

[HINT: press ESC until you see the main menu, enter 'analyze data' and then enter 'descriptive statistics'. If this is too complicated, just press Alt-T and type the next line and press ENTER, then F10 and then ENTER again.]

The next display shows what happened when I did this in October 1991:

FREQUENCIES PROV /FORMAT DFREQ.

The raw data or transformation pass is proceeding
408 cases are written to the uncompressed active file.

***** Memory allows a total of 9946 Values, accumulated across all Variables.
There also may be up to 1243 Value Labels for each Variable.

Page 6 SPSS/PC+ 10/25/91

PROV

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	3.00	180	44.1	44.1	44.1
	2.00	122	29.9	29.9	74.0
	1.00	106	26.0	26.0	100.0
	TOTAL	408	100.0	100.0	

[this tells us that 106 firms, or 26%, buy no raw materials from sources within their province, that 30% buy small amounts and 44% buy the majority of their inputs from provincial sources]
Valid Cases 408 Missing Cases 0

Page 7 SPSS/PC+ 10/25/91

This procedure was completed at 12:34:51

Notice that I neglected to put a value label on the new values 1, 2, and 3. Knowing how forgetful I am, I decided to go back and fix that up, so later I will know what 1 2 and 3 stand for. Here is the command line to put labels on each value. Can you figure out how to enter it?

VALUE LABELS PROV 1 '0%' 2 '1-50%' 3 '50-100%'.

And then I generated the frequencies once again, this time with the labels in place.

FREQUENCIES /VARIABLES PROV /FORMAT DFREQ.

***** Memory allows a total of 9946 Values, accumulated across all Variables.
There also may be up to 1243 Value Labels for each Variable.

Page 6 SPSS/PC+ 10/25/91

PROV

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
50-100%	3.00	180	44.1	44.1	44.1
1-50%	2.00	122	29.9	29.9	74.0
0%	1.00	106	26.0	26.0	100.0
	TOTAL	408	100.0	100.0	

Valid Cases 408 Missing Cases 0

Page 7 SPSS/PC+ 10/25/91

This procedure was completed at 12:34:51

Now let's see the percentage of input purchases made within the province by sector of manufacturing activity, V13. Another crosstabulation. Can you remember how to do it?]

CROSSTABS /TABLES V13 BY PROV /OPTIONS 5.

***** Given WORKSPACE allows for 7294 Cells with
2 Dimensions for CROSSTAB problem *****

Page 9 SPSS/PC+ 10/25/91
Crosstabulation: V13 Manuf-activity
By PROV

----- Page 1 of 4

PROV->	Count Tot Pct	1.00	2.00	3.00	Row Total
V13					
Handicraft	1 1.0	2 1.0	1 .5	5 2.6	8 4.2
Clothing-textile	2 2.6	5 2.6	7 3.6	7 3.6	19 9.9
Finished wood pr	3 .5	1 .5	3 1.6	14 7.3	18 9.4
Plastics	4 2.6	3 2.6	6 3.1		11 5.7
Column		41	68	83	192
(Continued) Total		21.4	35.4	43.2	100.0

Page 10 SPSS/PC+ 10/25/91
Crosstabulation: V13 Manuf-activity
By PROV

----- Page 2 of 4

PROV->	Count Tot Pct	1.00	2.00	3.00	Row Total
V13					
Metal products	5 3.6	7 3.6	13 6.8	9 4.7	29 15.1
Confectionery ba	6 1.0	2 1.0	4 2.1	26 13.5	32 16.7
Assembly elec-tr	7 .5	1 .5	1 .5	2 1.0	4 2.1
Assembly other	8 2.6		5 2.6		5 2.6
Column		41	68	83	192
(Continued) Total		21.4	35.4	43.2	100.0

Crosstabulation: V13 Manuf-activity
By PROV

- - - - Page 3 of 4

PROV->	Count Tot Pct	1.00	2.00	3.00	Row Total
V13					
9	4	5	2	11	
Breiving beverage	2.1	2.6	1.0	5.7	
10	1	1	1	3	
Machinery equipm	.5	.5	.5	1.6	
11	6	5	7	18	
Paper paper-prod	3.1	2.6	3.6	9.4	
12		5		5	
Leather-footwear		2.6		2.6	
Column	41	68	83	192	
(Continued) Total	21.4	35.4	43.2	100.0	

Crosstabulation: V13 Manuf-activity
By PROV

- - - - Page 4 of 4

PROV->	Count Tot Pct	1.00	2.00	3.00	Row Total
V13					
13	4	7	3	14	
Chemical pharmac	2.1	3.6	1.6	7.3	
14	3	5	7	15	
Other	1.6	2.6	3.6	7.8	
Column	41	68	83	192	
Total	21.4	35.4	43.2	100.0	

Number of Missing Observations = 216

Here is another idea to consider: perhaps you could look at averages by Sector, using the MEANS command. Try to set up that command line now.

D. Testing ideas using SPSS

Now let us combine some of these operations to try to learn more about the evolution of the Cameroonian economy. We will examine the age of each enterprise in the sample, examine their ages in relation to events in Cameroonian history, and then relate the age of the enterprise to sector of activity. We may learn which sectors have seen more or less effort at establishment.

First we will generate the table of simple frequencies of V4, the variable which directly relates to Question 4, "how long has the firm been operating." See if you can recall how to do this before looking at the following example.

FREQUENCIES /VARIABLES V4.

Value Label	YEARS Value	# FIRMS Frequency	Percent	Valid Percent	Cum Percent
	0	21	5.1	5.1	5.1
	1	30	7.4	7.4	12.5
	2	28	6.9	6.9	19.4
	3	41	10.0	10.0	29.4
	4	27	6.6	6.6	36.0
	5	22	5.4	5.4	41.4
	6	21	5.1	5.1	46.6
	7	17	4.2	4.2	50.7
	8	21	5.1	5.1	55.9
	9	12	2.9	2.9	58.8
<u>[CURRENT GOVERNMENT]</u>	10	30	7.4	7.4	66.2
	11	10	2.5	2.5	68.6
	12	10	2.5	2.5	71.1
	13	8	2.0	2.0	73.0
	14	9	2.2	2.2	75.2
	15	11	2.7	2.7	77.9
	16	3	.7	.7	78.7
	17	6	1.5	1.5	80.1
	18	13	3.2	3.2	83.3
	19	3	.7	.7	84.1
	20	13	3.2	3.2	87.3
	21	6	1.5	1.5	88.7
	22	6	1.5	1.5	90.2
	24	2	.5	.5	90.7
	25	4	1.0	1.0	91.7
	26	5	1.2	1.2	92.9
	27	1	.2	.2	93.1
	28	5	1.2	1.2	94.4
	30	6	1.5	1.5	95.8
<u>[1960/61 Independance]</u>	32	2	.5	.5	96.3
	35	1	.2	.2	96.6
	38	2	.5	.5	97.1
	39	2	.5	.5	97.5
	40	3	.7	.7	98.3
	41	3	.7	.7	99.0
	42	1	.2	.2	99.3
	50	2	.5	.5	99.8
	52	1	.2	.2	100.0
	TOTAL	408	100.0	100.0	

[This simple frequency table, which has been annotated with key events in Cameroonian history, can have meaning for analysts. But it is very long. taking into account these milestones in Cameroonian history, let regroup this variable transforming the data as follows:]

IF (V4 LE 5) AGE = 1. *[this, of course, creates a new variable, AGE, in which value 1 will include all firms aged 5 years and less]*

IF (V4 GT 5 AND V4 LE 10) AGE = 2.
 IF (V4 GT 10 AND V4 LE 15) AGE = 3.
 IF (V4 GT 15 AND V4 LE 20) AGE = 4.
 IF (V4 GT 20 AND V4 LE 25) AGE = 5.
 IF (V4 GT 25 AND V4 LE 30) AGE = 6.
 IF (V4 GT 30) AGE = 7.

[now to put in labels for values:]

VALUE LABELS AGE 1 '0-5 YRS' 3 '11-15 YRS' 5 '21-25 YRS' 7 '30 + YRS'.

[remember now to press F10 and then ENTER to start SPSS working]

FREQUENCIES /VARIABLES AGE.

The raw data or transformation pass is proceeding

408 cases are written to the uncompressed active file.

***** Memory allows a total of 4230 Values, accumulated across all Variables.

There also may be up to 529 Value Labels for each Variable.

[here is the first output: numbers of firms appearing in each value group]

Value Label	Value	<u>(#firms)</u> Frequency	Percent	Valid Percent	Cum Percent
0-5 YRS	1.00	169	41.4	41.4	41.4
	2.00	101	24.8	24.8	66.2
11-15 YRS	3.00	48	11.8	11.8	77.9
	4.00	38	9.3	9.3	87.3
21-25 YRS	5.00	18	4.4	4.4	91.7
	6.00	17	4.2	4.2	95.8
30 + YRS	7.00	17	4.2	4.2	100.0
TOTAL		408	100.0	100.0	
Valid Cases	408	Missing Cases	0		

[now we may be curious to see something about the formation of enterprises by sector of activity and the years in which they were formed. The CROSSTABS command is appropriate. We use AGE by variables V10A through V10F: consult the list of questions to see where we got these variables!]

CROSSTABS /TABLES AGE BY V10A TO V10F /OPTIONS 5.
[remember to use F10 and ENTER...]

Your results will appear in multiple tables on your screen.

Again, alternately, we can look at the average age of each type of enterprise, using the MEANS command. Calculating the means is a fast way to get a broad picture of the tendencies contained in a wide range of values. The MEANS command displays in a simple table the means, standard deviations, and group counts for a dependent variable within groups defined by one or more independent variables. In this case the dependent variable is the number of firms which purchase raw materials from provincial sources BY their sector of manufacturing activity. Do this now and compare the difference between the means and the data revealed above by crosstabulation.

MEANS /TABLES V19A BY V13.

Summaries of By levels of Variable	V19A V13 Value	Inputs-province Manuf-activity Label	Mean	Std Dev	Cases
For Entire Population					
V13	1	Handicraft	47.6250	42.0282	192
V13	2	Clothing-textiles	58.7500	47.9397	8
V13	3	Finished wood produc	39.6316	41.5240	19
V13	4	Plastics	76.6667	30.2927	18
V13	5	Metal products	9.7273	14.9271	11
V13	6	Confectionery baking	36.9655	39.1412	29
V13	7	Assembly elec-trical	84.0625	33.3949	32
V13	8	Assembly other	46.5000	50.9477	4
V13	9	Breving beverage pro	21.0000	18.8414	5
V13	10	Machinery equipment	23.1818	34.6607	11
V13	11	Paper paper-products	38.3333	46.4579	3
V13	12	Leather-footwear	40.2778	41.9549	18
V13	13	Chemical pharmaceuti	22.2000	18.9526	5
V13	14	Other	28.2857	30.4440	14
			51.9333	43.8707	15
Total Cases =		408			
Missing Cases =		216 OR 52.9 PCT.			

E. Three-way Contingency: thinness of data

What about looking into the relation between age of the enterprise and sector of activity in terms of whether the enterprise is formal or informal? We might learn more about which points of entry into the economy are more available to new operators. We might learn more about which sectors are contributing relevant shares of revenues to the state.

We start as before, by regrouping Variable 4 into easier-to-use categories defined by five year intervals of existence:

```
IF (v4 le 5) age = 1.
IF (V4 GT 5 AND V4 LE 10) AGE = 2.
IF (V4 GT 10 AND V4 LE 15) AGE = 3.
IF (V4 GT 15 AND V4 LE 20) AGE = 4.
IF (V4 GT 20 AND V4 LE 25) AGE = 5.
IF (V4 GT 25 AND V4 LE 30) AGE = 6.
IF (V4 GT 30) AGE =7.
```

[then we provide useful labels so we can remember what these values stand for:]

```
VALUE LABELS AGE 1 '0-5yrs' 3 '11-15yrs' 5 '21-25yrs' 7 '30yrs +'.
```

[and we run a cross tabulation of AGE of enterprise BY sector of activity (Question 10, choices A through F) controlling for formality, V1:]

```
CROSSTABS /VARIABLES AGE BY V10A TO V10F BY V1/ OPTIONS 5.
The raw data or transformation pass is proceeding
408 cases are written to the uncompressed active file.
```

```
***** Given WORKSPACE allows for 2449 Cells with
3 Dimensions for CROSSTAB problem *****
```

[after a good dealing of machine noise and calculation, a BEEP and a MORE message in the upper right hand corner, the results come forth. We want to notice, however, that when we crosstabulate by three variables, by necessity fewer and fewer cases will match all three conditions. This means that meaningful generalizations from the results become less and less safe.]

You will see from the first set of frames, Agriculture, that, as far as the informal sector is concerned, this type of analysis reveals only that we did not survey any informal/agriculture sector enterprises. Continue through all the frames and make notes of possible areas of interest.

Crosstabulation: AGE
 By V10A Agriculture
 Controlling for V1 Sampling source
 = 1 Formal registry
 - - - - Page 1 of 3

V10A→	Count Tot Pct	Yes		No		Row Total
		1	2	1	2	
AGE						
0-5yrs	1.00	3	89			92
		1.0	29.1			30.1
	2.00	5	75			80
		1.6	24.5			26.1
11-15yrs	3.00	8	38			46
		2.6	12.4			15.0
Column		25	281			306
(Continued) Total		8.2	91.8			100.0

Page 16 SPSS/PC+ 10/25/91

Crosstabulation: AGE
 By V10A Agriculture
 Controlling for V1 Sampling source
 = 1 Formal registry
 - - - - Page 2 of 3

V10A→	Count Tot Pct	Yes		No		Row Total
		1	2	1	2	
AGE						
	4.00	4	32			36
		1.3	10.5			11.8
21-25yrs	5.00	1	17			18
		.3	5.6			5.9
	6.00		17			17
			5.6			5.6
Column		25	281			306
(Continued) Total		8.2	91.8			100.0

Page 17 SPSS/PC+ 10/25/91

Crosstabulation: AGE
 By V10A Agriculture
 Controlling for V1 Sampling source
 = 1 Formal registry
 - - - - Page 3 of 3

V10A→	Count Tot Pct	Yes		No		Row Total
		1	2	1	2	
AGE						
30ys +	7.00	4	13			17
		1.3	4.2			5.6
Column		25	281			306
Total		8.2	91.8			100.0

Crosstabulation: AGE
 By V10A Agriculture
 Controlling for V1 Sampling source
 = 2 Informal sampling
 - - - - Page 1 of 2

V10A->	Count Tot Pct	No 2	Row Total
AGE			
0-5yrs	1.00 77	77	77
		75.5	75.5
	2.00	21	21
		20.6	20.6
11-15yrs	3.00	2	2
		2.0	2.0
Column		102	102
(Continued) Total		100.0	100.0

Crosstabulation: AGE
 By V10A Agriculture
 Controlling for V1 Sampling source
 = 2 Informal sampling
 - - - - Page 2 of 2

V10A->	Count Tot Pct	No 2	Row Total
AGE			
	4.00	2	2
		2.0	2.0
Column		102	102
Total		100.0	100.0

Number of Missing Observations = 0

[this next table is very interesting; notice the relatively large number of informal enterprises which have recently entered manufacturing!]

Crosstabulation: AGE
 By V10C Manufacturing
 Controlling for V1 Sampling source
 = 2 Informal sampling
 - - - - Page 1 of 2

V10C->	Count Tot Pct	Yes 1	No 2	Row Total
AGE				
0-5yrs	1.00	23	54	77
		22.5	52.9	75.5
	2.00	6	15	21
		5.9	14.7	20.6
11-15yrs	3.00		2	2
			2.0	2.0
Column		30	72	102
(Continued) Total		29.4	70.6	100.0

Crosstabulation: AGE
 By V10C Manufacturing
 Controlling for V1 Sampling source

= 2 Informal sampling
 - - - - Page 2 of 2

V10C→	Count Tot Pct	Yes 1	No 2	Row Total
AGE	4.00	1 1.0	1 1.0	2 2.0
Column		30	72	102
Total		29.4	70.6	100.0

Number of Missing Observations = 0

Crosstabulation: AGE
 By V10D Commerce
 Controlling for V1 Sampling source

= 1 Formal registry
 - - - - Page 1 of 3

V10D→	Count Tot Pct	Yes 1	No 2	Row Total
AGE	1.00	49 16.0	43 14.1	92 30.1
0-5yrs	2.00	46 15.0	34 11.1	80 26.1
11-15yrs	3.00	19 6.2	27 8.8	46 15.0
Column		148	158	306
(Continued) Total		48.4	51.6	100.0

Crosstabulation: AGE
 By V10D Commerce
 Controlling for V1 Sampling source

= 1 Formal registry
 - - - - Page 2 of 3

V10D→	Count Tot Pct	Yes 1	No 2	Row Total
AGE	4.00	9 2.9	27 8.8	36 11.8
21-25yrs	5.00	11 3.6	7 2.3	18 5.9
6.00	7 2.3	10 3.3	17 5.6	
Column		148	158	306
(Continued) Total		48.4	51.6	100.0

Crosstabulation: AGE
 By V10D Commerce
 Controlling for V1 Sampling source

= 1 Formal registry
 - - - - Page 3 of 3

V10D→	Count Tot Pct	Yes 1	No 2	Row Total
AGE	7.00	7 2.3	10 3.3	17 5.6
30yrs +	Column	148	158	306
Total		48.4	51.6	100.0

Page 33 SPSS/PC+ 10/25/91
Crosstabulation: AGE
By V10D Commerce
Controlling for V1 Sampling source
= 2 Informal sampling
- - - - Page 1 of 2

V10D→	Count Tot Pct	Yes 1	No 2	Row Total
AGE				
0-5yrs	1.00 38.2	39	38	77
	2.00 11.8	12	9	21
	3.00 1.0	1	1	2
11-15yrs				
Column		54	48	102
(Continued) Total		52.9	47.1	100.0

Page 34 SPSS/PC+ 10/25/91
Crosstabulation: AGE
By V10D Commerce
Controlling for V1 Sampling source
= 2 Informal sampling
- - - - Page 2 of 2

V10D→	Count Tot Pct	Yes 1	No 2	Row Total
AGE				
	4.00 2.0	2		2
Column		54	48	102
Total		52.9	47.1	100.0
Number of Missing Observations =				0

There are many more screens, of course, in this series. Take some time to look at each one and determine the value of the data for your purposes. How would you have done this crosstabulation differently? What would be the value of using the MEANS command?

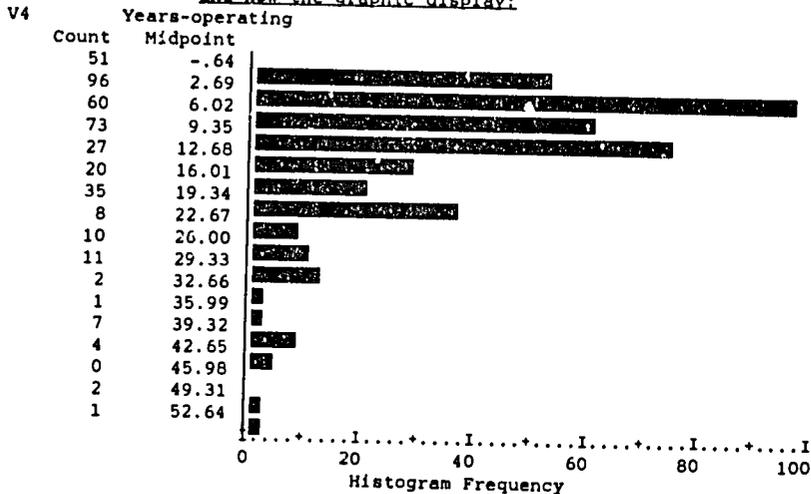
The SPSS/PC+ Histogram command is also found under the FREQUENCIES selection under the 'analyze data' choice in the main menu. The command line first produces the table of numerical values, in this case, the number of enterprises surveyed by number of years they have been operating:

FREQUENCIES /VARIABLES V4 /HISTOGRAM.

V4 Years-operating

Value Label	Value	Frequency	Percent	Valid Percent	Cum Percent
	0	21	5.1	5.1	5.1
	1	30	7.4	7.4	12.5
	2	28	6.9	6.9	19.4
	3	41	10.0	10.0	29.4
	4	27	6.6	6.6	36.0
	5	22	5.4	5.4	41.4
	6	21	5.1	5.1	46.6
	7	17	4.2	4.2	50.7
	8	21	5.1	5.1	55.9
	9	12	2.9	2.9	58.8
	10	30	7.4	7.4	66.2
	11	10	2.5	2.5	68.6
	12	10	2.5	2.5	71.1
	13	8	2.0	2.0	73.0
	14	9	2.2	2.2	75.2
	15	11	2.7	2.7	77.9
	16	3	.7	.7	78.7
	17	6	1.5	1.5	80.1
	18	13	3.2	3.2	83.3
	19	3	.7	.7	84.1
	20	13	3.2	3.2	87.3
	21	6	1.5	1.5	88.7
	22	6	1.5	1.5	90.2
	24	2	.5	.5	90.7
	25	4	1.0	1.0	91.7
	26	5	1.2	1.2	92.9
	27	1	.2	.2	93.1
	28	5	1.2	1.2	94.4
	30	6	1.5	1.5	95.8
	32	2	.5	.5	96.3
	35	1	.2	.2	96.6
	38	2	.5	.5	97.1
	39	2	.5	.5	97.5
	40	3	.7	.7	98.3
	41	3	.7	.7	99.0
	42	1	.2	.2	99.3
	50	2	.5	.5	99.8
	52	1	.2	.2	100.0
		TOTAL	408	100.0	100.0

and now the graphic display:



We can also discover the Means for variables, controlling for some condition. In the following case we are asking for a calculation of the Mean, Minimum, Maximum, Median, Variance, etc for the age of the enterprise, controlled for formality or informality. The command line concludes with a requirement that SPSS/PC+ display the results both in a table and as a histogram:

EXAMINE /VARIABLES V4 BY V1 /PLOT HISTOGRAM.

```

V4
Valid cases:      408.0  Missing cases:      .0  Percent missing:      .0

Mean      10.1642  Std Err      .4823  Min      .0000  Skewness      1.6039
Median    7.0000  Variance     94.9140  Max      52.0000  S E Skew      .1208
5% Trim   9.1367  Std Dev      9.7424  Range    52.0000  Kurtosis     2.7565
                                IQR      11.0000  S E Kurt     .2411

```

```

Frequency  Bin Center
240.00     5  *****
103.00    15  *****
 42.00    25  *****
  6.00    35  *
 17.00  Extremes  ***

```

Bin width : 10 The "bin width" is ten years, the bin center is therefore mid-way, or five years.
Each star: 5 case(s)

```

V4
By V1      1      Formal registry
Valid cases: 306.0  Missing cases:  .0  Percent missing:  .0

Mean      12.2745  Std Err      .5830  Min      .0000  Skewness      1.3722
Median    10.0000  Variance    104.0162  Max      52.0000  S E Skew      .1393
5% Trim   11.3003  Std Dev     10.1988  Range    52.0000  Kurtosis     1.8794
                                IQR      14.0000  S E Kurt     .2778

```

```

V4
By V1      1      Formal registry
Frequency  Bin Center
145.00     5  *****
 97.00    15  *****
 41.00    25  *****
 11.00    35  ****
 12.00  Extremes  ****

```

Bin width : 10
Each star: 3 case(s)

```

V4
By V1      2      Informal sampling
Valid cases: 102.0  Missing cases:  .0  Percent missing:  .0

Mean      3.8333  Std Err      .3757  Min      .0000  Skewness      1.7274
Median    3.0000  Variance    14.3977  Max      20.0000  S E Skew      .2391
5% Trim   3.4314  Std Dev      3.7944  Range    20.0000  Kurtosis     3.9871
                                IQR      4.2500  S E Kurt     .4738

```

```

Frequency  Bin Center
 95.00     5  *****
  3.00    15  *
  4.00  Extremes  **

```

Bin width : 10
Each star: 2 case(s)

SCATTERPLOT SYMBOLS FOR MULTIPLE CASES:

Frequencies and symbols used
(not applicable for control or overlay plots)

# - Symb	# - Symb	# - Symb	# - Symb
1 - 1	11 - B	21 - L	31 - V
2 - 2	12 - C	22 - M	32 - W
3 - 3	13 - D	23 - N	33 - X
4 - 4	14 - E	24 - O	34 - Y
5 - 5	15 - F	25 - P	35 - Z
6 - 6	16 - G	26 - Q	36 - *
7 - 7	17 - H	27 - R	
8 - 8	18 - I	28 - S	
9 - 9	19 - J	29 - T	
10 - A	20 - K	30 - U	

PLOT /PLOT V4 WITH V80.
PLOT requires 6256 BYTES of workspace for execution.

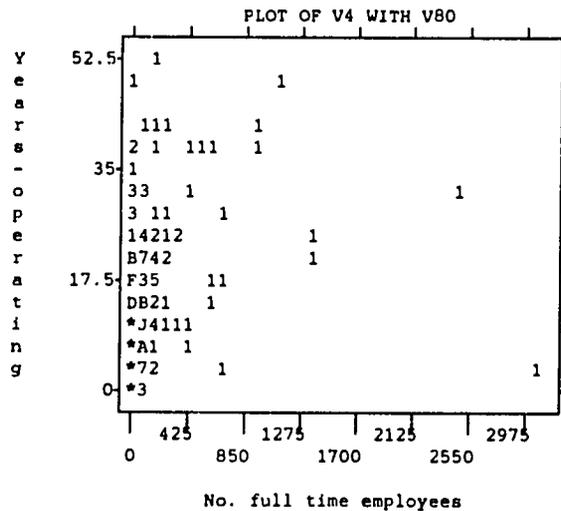
-----Page 20
SPSS/PC+ 10/18/91

***** P L O T *****

Data Information

408 unweighted cases accepted.

-----Page 21
SPSS/PC+ 10/18/91



Everything is jammed to the left, making it difficult to understand. And notice that only a few cases lie beyond 1275. Let's limit the value of V80 to less than 1200 to see if the picture becomes clearer.

-----Page 22
SPSS/PC+ 10/18/91

408 cases plotted.

B. FASTGRAF

1. Setting up the computer and printer

First get to the C>SPSS line, type GBSETUP and press ENTER. Then enter the SCREEN option and set the screen and graphics options to match what you have. Then enter the PRINTER option and select the type of printer you will be using. Once this is set up, it will hold for that computer and printer for all sessions. If you install SPSS in another computer, you may have to run the GBSETUP command again and re-configure.

2. Using FASTGRAF.

FASTGRAF, a separate program module created for SPSS by New England Software, will generate barcharts, line charts and pie charts from data which you have already generated through an SPSS command, such as FREQUENCIES.

First generate some data to graph

Second, move to the main menu with the ESC key

Third, move the cursor bar to 'graph data'

Fourth, got down to FASTGRAF, press ENTER, press F10 and ENTER

Fifth, the Graph-in-the-Box menu will appear. Press Alt-G to activate FASTGRAF.

Sixth, you will see your last table of data appear on the screen. Using the arrow keys to control the cursor, place the cursor over the first digit that you want to capture, press ENTER and move the cursor down to the last digit you want to capture and press ENTER again.

Seventh, you will see a menu appear at the bottom of the screen, with the cursor, select Show and your graph will appear. Hit the escape and select Display, and you can select a different type of graph, using the SPACE BAR. You can also type in a title for the graph, and labels for the x and y axes. Hit the escape again and select Data and you can fill in the labels for each value.

Eighth, you can print the graph by hitting escape and selecting Printer.

Ninth, you can end the process by selecting Quit, and pressing F10, which should return you to the main SPSS menu.

V. REGRESSIONS

First let's look at a picture of the data, to see what sort of distribution it may have. Let's look at the AGE of the firms with the SIZE of firms as expressed by the number of full time employees. Once we have a better feel for the relationship among the variables, we can decide which ones to analyze in a regression.

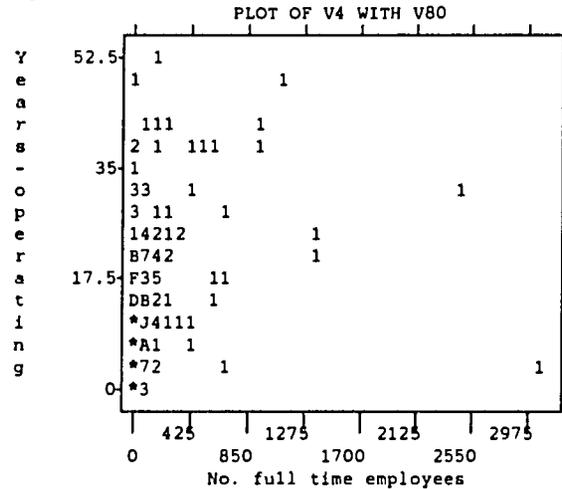
PLOT /PLOT V4 WITH V80.

PLOT requires 6256 BYTES of workspace for execution.

 Page 20 SPSS/PC+ 10/18/91
 ***** P L O T *****

Data Information
 408 unweighted cases accepted.

 Page 21 SPSS/PC+ 10/18/91



Everything is jammed to the left, making it difficult to understand. And notice that only a few cases lie beyond 1275. Let's limit the value of V80 to less than 1200 to see if the picture becomes clearer.

 Page 22 SPSS/PC+ 10/18/91

408 cases plotted.

SCATTERPLOT SYMBOLS FOR MULTIPLE CASES:

Frequencies and symbols used
 (not applicable for control or overlay plots)

# - Symb	# - Symb	# - Symb	# - Symb
1 - 1	11 - B	21 - L	31 - V
2 - 2	12 - C	22 - M	32 - W
3 - 3	13 - D	23 - N	33 - X
4 - 4	14 - E	24 - O	34 - Y
5 - 5	15 - F	25 - P	35 - Z
6 - 6	16 - G	26 - Q	36 - *
7 - 7	17 - H	27 - R	
8 - 8	18 - I	28 - S	
9 - 9	19 - J	29 - T	
10 - A	20 - K	30 - U	

Clear we have to place an upper limit on number of employees to see what is happening on the left side of the plot:

PROCESS IF (V80 LE 1200).

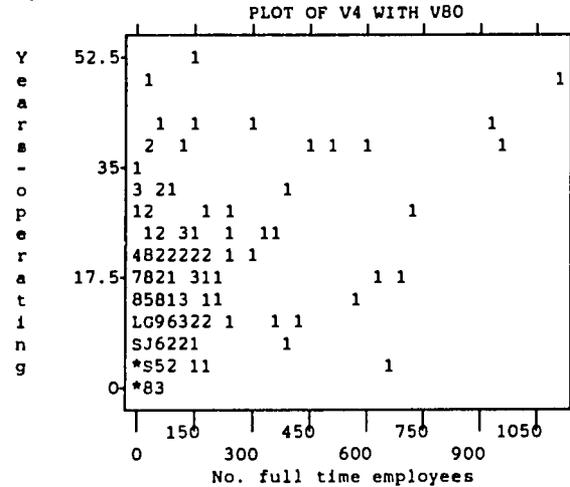
We are telling SPSS/PC+ to process variable 80 only if the value is less than or equal to 1200 employees.

PLOT /PLOT V4 WITH V80.
 PLOT requires 6256 BYTES of workspace for execution.

 Page 26 SPSS/PC+ 10/18/91
 ***** P L O T *****

Data Information
 404 unweighted cases accepted. [we lost only 4 cases!]

 Page 27 SPSS/PC+ 10/18/91



[there is something of a 45 degree angle line, but there is still a big cluster of firms centered at 17.5 years and less than 150 employees. Perhaps there is no linear relation between years in business and size of firm.]

[let's limit our data more - to firms of 750 or fewer workers]

 Page 28 SPSS/PC+ 10/18/91
 404 cases plotted.

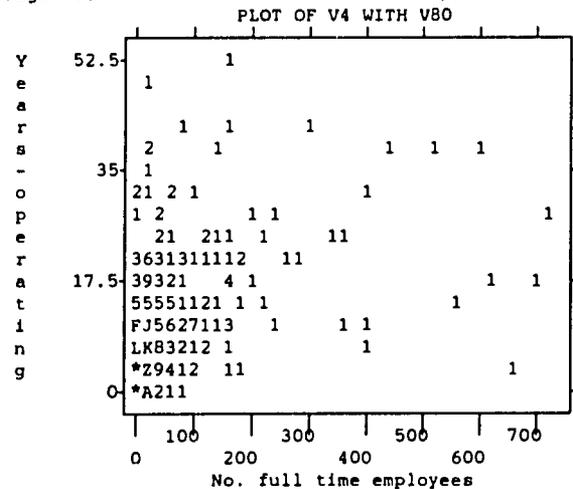
PROCESS IF (V80 LE 750). [again, here is the command limiting V80 to 750 or less]

PLOT /PLOT V4 WITH V80.
 PLOT requires 6256 BYTES of workspace for execution.

 Page 32 SPSS/PC+ 10/18/91
 ***** P L O T *****

Data Information
 401 unweighted cases accepted. [we lost only 7 cases]

 Page 33 SPSS/PC+ 10/18/91



[still no clear linear relation!]

 Page 34 SPSS/PC+ 10/18/91
 401 cases plotted.

REGRESSION /VARIABLES v80 V4 /DEPENDENT v80 /method ENTER.

***** MULTIPLE REGRESSION *****
 Listwise Deletion of Missing Data
 Equation Number 1 Dependent Variable.. V80 No. full time employees
 Beginning Block Number 1. Method: Enter

Page 39 SPSS/PC+ 10/18/91
 ***** MULTIPLE REGRESSION *****
 Equation Number 1 Dependent Variable.. V80 No. full time employees
 Variable(s) Entered on Step Number
 1.. V4 Years-operating
 Multiple R .32770
 R Square .10739
 Adjusted R Square .10519
 Standard Error 234.24641
 Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	2680166.50802	2680166.50802
Residual	406	22277779.63904	54871.37842

F = 48.84453 Signif F = .0000

Page 40 SPSS/PC+ 10/18/91
 ***** MULTIPLE REGRESSION *****
 Equation Number 1 Dependent Variable.. V80 No. full time employees

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
V4	8.32949	1.19182	.32770	6.989	.0000
(Constant)	-5.38335	16.77008		-.321	.7484

End Block Number 1 All requested variables entered.

PROCESS IF (V80 LE 750).

Page 43 SPSS/PC+ 10/18/91
 regression /variables v80 v4 /dependent v80 /method enter.
 ***** MULTIPLE REGRESSION *****
 Listwise Deletion of Missing Data
 Equation Number 1 Dependent Variable.. V80 No. full time employees
 Beginning Block Number 1. Method: Enter

Page 44 SPSS/PC+ 10/18/91
 ***** MULTIPLE REGRESSION *****
 Equation Number 1 Dependent Variable.. V80 No. full time employees
 Variable(s) Entered on Step Number
 1.. V4 Years-operating
 Multiple R .42855
 R Square .18366
 Adjusted R Square .18161
 Standard Error 92.74452
 Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	1	772123.66276	772123.66276
Residual	399	3432017.21006	8601.54689

F = 89.76567 Signif F = .0000

Page 45 SPSS/PC+ 10/18/91
 ***** MULTIPLE REGRESSION *****
 Equation Number 1 Dependent Variable.. V80 No. full time employees

Variables in the Equation

Variable	B	SE B	Beta	T	Sig T
V4	4.74343	.50065	.42855	9.474	.0000
(Constant)	5.99960	6.75271		.888	.3748

End Block Number 1 All requested variables entered.

CORRELATIONS /VARIABLES V4 V21 V22A V22B V22C V22D V22E V22F V22G V23 V27.

Correlations:	V4	V21	V22A	V22B	V22C	V22D
V4	1.0000	.0747	.3916**	-.0110	-.0313	-.0624
V21	.0747	1.0000	-.0049	-.0135	-.0162	-.0163
V22A	.3916**	-.0049	1.0000	-.1253	-.1377	-.0858
V22B	-.0110	-.0135	-.1253	1.0000	-.0255	-.0257
V22C	-.0313	-.0162	-.1377	-.0255	1.0000	-.0308
V22D	-.0624	-.0163	-.0858	-.0257	-.0308	1.0000
V22E	-.0028	-.0266	-.2945**	-.0420	.0180	-.0508
V22F	-.0969	-.0206	-.2274*	-.0324	-.0389	-.0392
V22G	-.3174**	-.0579	-.5810**	-.0913	-.0906	-.1104
V23	-.3186**	-.2748*	-.1251	.0490	-.0255	.0593
V27	.3613**	.0850	.1619	.0916	.0846	.0276

N of cases: 118 1-tailed Signif: * - .01 ** - .001
 " . " is printed if a coefficient cannot be computed

Page 5 SPSS/PC+ 10/18/91

Correlations:	V22E	V22F	V22G	V23	V27
V4	-.0028	-.0969	-.3174**	-.3186**	.3613**
V21	-.0266	-.0206	-.0579	-.2748*	.0850
V22A	-.2945**	-.2274*	-.5810**	-.1251	.1619
V22B	-.0420	-.0324	-.0913	.0490	.0916
V22C	.0180	-.0389	-.0906	-.0255	.0846
V22D	-.0508	-.0392	-.1104	.0593	.0276
V22E	1.0000	-.0641	-.1805	.0969	.0160
V22F	-.0641	1.0000	-.1094	.0748	-.2243*
V22G	-.1805	-.1094	1.0000	.1415	-.1390
V23	.0969	.0748	.1415	1.0000	-.2360**
V27	.0160	-.2243*	-.1390	-.2360**	1.0000

N of cases: 118 1-tailed Signif: * - .01 ** - .001
 " . " is printed if a coefficient cannot be computed

CORRELATIONS /VARIABLES V4 V21 V22A V22B V22C V22D V22E V22F V22G V80 V27.

Correlations:	V4	V21	V22A	V22B	V22C	V22D
V4	1.0000	.0747	.3916**	-.0110	-.0313	-.0624
V21	.0747	1.0000	-.0049	-.0135	-.0162	-.0163
V22A	.3916**	-.0049	1.0000	-.1253	-.1377	-.0858
V22B	-.0110	-.0135	-.1253	1.0000	-.0255	-.0257
V22C	-.0313	-.0162	-.1377	-.0255	1.0000	-.0308
V22D	-.0624	-.0163	-.0858	-.0257	-.0308	1.0000
V22E	-.0028	-.0266	-.2945**	-.0420	.0180	-.0508
V22F	-.0969	-.0206	-.2274*	-.0324	-.0389	-.0392
V22G	-.3174**	-.0579	-.5810**	-.0913	-.0906	-.1104
V80	.4666**	.0768	.1178	.2984**	.0923	-.0790
V27	.3613**	.0850	.1619	.0916	.0846	.0276

N of cases: 118 1-tailed Signif: * - .01 ** - .001
 " . " is printed if a coefficient cannot be computed

Page 8 SPSS/PC+ 10/18/91

Correlations:	V22E	V22F	V22G	V80	V27
V4	-.0028	-.0969	-.3174**	.4666**	.3613**
V21	-.0266	-.0206	-.0579	.0768	.0850
V22A	-.2945**	-.2274*	-.5810**	.1178	.1619
V22B	-.0420	-.0324	-.0913	.2984**	.0916
V22C	.0180	-.0389	-.0906	.0923	.0846
V22D	-.0508	-.0392	-.1104	-.0780	.0276
V22E	1.0000	-.0641	-.1805	-.1026	.0160
V22F	-.0641	1.0000	-.1094	-.1134	-.2243*
V22G	-.1805	-.1094	1.0000	-.1155	-.1390
V80	-.1026	-.1134	-.1155	1.0000	.3621**
V27	.0160	-.2243*	-.1390	.3621**	1.0000

N of cases: 118 1-tailed Signif: * - .01 ** - .001
 " . " is printed if a coefficient cannot be computed

Having looked at the correlations, now we are ready to run a regression on a few of them: V4 will be the dependent, and V22a, V80, V27 and V21 the independent.

REGRESSION /VARIABLES v4 v22a v80 v27 v21 /DEPENDENT v4 /METHOD ENTER.

***** MULTIPLE REGRESSION *****
Listwise Deletion of Missing Data
Equation Number 1 Dependent Variable.. V4 Years-operating
Beginning Block Number 1. Method: Enter

Page 11 SPSS/PC+ 10/18/91

***** MULTIPLE REGRESSION *****
Equation Number 1 Dependent Variable.. V4 Years-operating
Variable(s) Entered on Step Number
1.. V21 Nationality-owner
2.. V22A French-owner
3.. V80 No. full time employees
4.. V27 CA

Page 12 SPSS/PC+ 10/18/91

***** MULTIPLE REGRESSION *****
Equation Number 1 Dependent Variable.. V4 Years-operating
Multiple R .60055
R Square .36066
Adjusted R Square .33803
Standard Error 9.34407
Analysis of Variance
DF Sum of Squares Mean Square
Regression 4 5565.65439 1391.41360
Residual 113 9866.21849 87.31167
F * 15.93617 Signif F = .0000

Page 13 SPSS/PC+ 10/18/91

***** MULTIPLE REGRESSION *****
Equation Number 1 Dependent Variable.. V4 Years-operating
----- Variables in the Equation -----
Variable B SE B Beta T Sig T
V21 4.18651 9.43190 .03356 .444 .6580
V22A .08269 .01970 .32071 4.198 .0001
V80 .01605 3.58216E-03 .36284 4.481 .0000
V27 1.30807 .60919 .17509 2.147 .0339
(Constant) -3.43464 18.93574 -.181 .8564
End Block Number 1 All requested variables entered.

RECODE V21 (1=1) (2=0) (3=0).

Page 16 SPSS/PC+ 10/18/91

RECODE V10A 1\ V10F (1=1) (2=0).

regression /variables v4 v10a to v10f v21 v22a v27 v80 /dependent v4 /method
 The raw data or transformation pass is proceeding
 408 cases are written to the uncompressed active file.
 enter.

 Page 18 SPSS/PC+ 10/18/91
 * * * * * M U L T I P L E R E G R E S S I O N * * * * *

Listwise Deletion of Missing Data
 Equation Number 1 Dependent Variable.. V4 Years-operating
 The following variables are constants or have missing correlations:
 V21
 They will be deleted from the analysis.

 Page 19 SPSS/PC+ 10/18/91
 * * * * * M U L T I P L E R E G R E S S I O N * * * * *

Equation Number 1 Dependent Variable.. V4 Years-operating
 Beginning Block Number 1. Method: Enter

 Page 20 SPSS/PC+ 10/18/91
 * * * * * M U L T I P L E R E G R E S S I O N * * * * *

Equation Number 1 Dependent Variable.. V4 Years-operating
 Variable(s) Entered on Step Number
 1.. V80 No. full time employees
 2.. V10E Other services
 3.. V10F Mining
 4.. V22A French-owner
 5.. V10A Agriculture
 6.. V10D Commerce
 7.. V27 CA
 8.. V10B Agroprocessing
 9.. V10C Manufacturing

 Page 21 SPSS/PC+ 10/18/91
 * * * * * M U L T I P L E R E G R E S S I O N * * * * *

Equation Number 1 Dependent Variable.. V4 Years-operating
 Multiple R .63388
 R Square .40181
 Adjusted R Square .35196
 Standard Error 9.24525
 Analysis of Variance

	DF	Sum of Squares	Mean Square
Regression	9	6200.60909	688.95657
Residual	108	9231.26380	85.47466
F =	8.06036	Signif F = .0000	

 Page 22 SPSS/PC+ 10/18/91
 * * * * * M U L T I P L E R E G R E S S I O N * * * * *

Equation Number 1 Dependent Variable.. V4 Years-operating

----- Variables in the Equation -----

Variable	B	SE B	Beta	T	Sig T
V80	.01725	3.81429E-03	.38980	4.521	.0000
V10E	6.38646	2.61250	.23147	2.445	.0161
V10F	7.05202	5.77608	.09707	1.221	.2248
V22A	.07890	.01982	.30601	3.980	.0001
V10A	3.86407	3.54799	.09824	1.089	.2785
V10D	2.34159	1.90048	.10090	1.232	.2206
V27	1.03969	.62960	.13916	1.651	.1016
V10B	-.24767	2.77846	-8.125E-03	-.089	.9291
V10C	4.43716	2.11695	.19389	2.096	.0384
(Constant)	.91652	3.01033		.304	.7614

 End Block Number 1 All requested variables entered.

VI. Troubleshooting

THE COMPUTER SAYS IT CAN'T DO WHAT I WANT

1. make certain there is at least 2,000,000 bytes free on the hard disk; complex calculations make heavy demands on memory space.
2. if you get ERROR lines, you have not typed in the command in the form that SPSS can understand. Check that there are spaces between command elements (remember the problem with LE50 earlier? Make certain that you have put each command element in the right order.

NOTHING HAPPENS!

1. Did you remember to press F10 and ENTER?
2. Did you remember to end each command line with a period (.)?
3. Did the computer BEEP at you and show MORE in the upper right hand corner? If so, just press the SPACE BAR to get it to move along.
4. Did you start out by loading all the variables into the active memory? If not, there is nothing to process.

THE WHOLE THING TAKES A LONG TIME!

1. Indeed, it certainly seems that way in the middle of a session.
2. Computers run at different speeds, from 2Mhz (an old Tandy) to as much as 40 or 50Mhz. The speed depends on the CPU chip. an 8086 is slower than an 8088, which is slower than an 80286, which is slower than an 80386, which is slower than an 80486. Speed also depends on how much RAM memory you have free.
3. Another issue is the complexity of the task which you set SPSS to carry out. A simple FREQUENCIES table goes faster than a complex regrouping followed by a labeling and crosstabulation.

THE GRAPHICS THING WON'T WORK!

1. SPSS requires additional software modules in many cases. We can use PLOT and FASTGRAF with the modules we have.
2. Did you remember to run the set-up utility before starting? If not, get out of SPSS, get to the c>SPSS line, and type GBSETUP and press ENTER.

VII. How to Import the database into other software

The database contained in SPSS software can be transformed to be used by other software such as Lotus 1,2,3 or dBase. This can be useful in preparing spreadsheets, or in manipulating the data in other ways. The SPSS/PC+ command EXPORT is the mechanism to accomplish this.

SPSS/PC+ 's EXPORT command produces a portable ASCII data file and dictionary. You can upload EXPORT files to a mainframe computer using the KERMIT communications protocol, provided that KERMIT is installed on both the PC (clone) and the receiving computer. KERMIT is usually one of the options available in most communications software, such as CrossTalk, ProComm, etc.

A. Preparing your data

EXPORT is limited to a maximum of 200 variables on a portable file. As the Cameroon data base has more than 364 variables, we will have to divide the database into two parts. You may recall that the database originally came in seven OUTPUT.SYS files. For safety's sake, let us experiment using just one .SYS file.

Here is a list of the .SYS files and the variables each contains:

1. OUTPUT1.SYS: QUEST., V1-V16j, V94, V95
2. OUTPUT2.SYS: V17a-V20e
3. OUTPUT3.SYS: V21-V32j
4. OUTPUT4.SYS: V33a-V42
5. OUTPUT5.SYS: V43a-V54e
6. OUTPUT6.SYS: V55a-V63
7. OUTPUT7.SYS: V64-V77h

B. Exporting

EXPORT is a data transformation command. The command line syntax is as follows:

```
EXPORT OUTFILE='NEWDATA1.POR'.
```

The suffix *.POR tells SPSS that you are going to create a portable file, that is, one that will reside outside the *.SYS files. The OUTFILE subcommand specifies the new name of the portable file, in this case, NEWDATA1.POR. We can also direct this OUTFILE to the floppy disc drive. If it is the A drive, then call the file 'A:\NEWDATA1.POR'

Now we have to limit the number of variables. Let's say that you are interested for the purposes of your spreadsheet analysis

only in the data contained in variables 33 through 41. We can specify which variables should be exported:

```
EXPORT OUTFILE='NEWDATA1.POR' /KEEP V33 TO V41.
```

The variables not named in the KEEP subcommand will not be exported. That solves several problems: the 200 variable limitation of the EXPORT command, filling up disc and memory space needlessly.

To make certain you remember which variables to exported to which *.POR file, add the subcommand /MAP to the end of the command line, but before the period (.).

```
EXPORT OUTFILE='NEWDATA1.POR' /KEEP V33 TO V41 /MAP.
```

C. Importing

Now you have created an ASCII delimited file. Open up Lotus or dBase and import the database and have fun!

VIII. LOTUS 123, and EXCEL FOR WINDOWS

In addition to the SPSS/PC+ database, data on the Cameroon private sector are also available in the form of Lotus 123 worksheets, in the WK1 format. These data are available for all transformations, recalculation and graphing. As they are provided in Lotus WK1, they may also be imported into Microsoft Excel for Windows, and other spreadsheet programs.

TRANSLATE FROM 'xxxx.wk1' RANGE A1...22

	Var 1	Var 2	Var 3
Case 1			
Case 2			

IX. Definitions: a glossary

ACTIVE FILE - the file defined in a session by a DATA LIST command or called by a GET or IMPORT or INCLUDE command. The active file consists of the data and a dictionary and is the file that you modify using transformations, and analyze using any of the procedures.

ACTIVE MEMORY - the RAM memory (and sometimes disc memory) that is available to you during a session.

CASE, or OBSERVATION - the basic unit of analysis for which measurements are taken. In this instance, a CASE represents an interviewed enterprise.

CATEGORICAL VARIABLE - a variable that has a limited number of values that form categories of cases: Male or Female, Yes or No, etc.

COMMAND - a specific instruction that controls the execution of SPSS/PC+

COMMAND FILE - a file that contains SPSS/PC+ commands and sometimes also includes data. We use the INCLUDE command to process our command files, such as CAMEROON.LOG

COMMAND LINE - the line of specific instructions for one SPSS/PC+ command. It may take more than one physical line on the screen to enter all of the instructions for a single command.

COMPOSITE VARIABLE - a variable we create by combining or manipulating simple variables: Enterprises owned by females.

CONSTANT - something the value of which does not change; we measure variables quite often against a constant.

CONTINGENCY TABLE - a table containing the joint frequency distribution of two or more variables, invoked by the CROSSTABS command

CONTINUOUS VARIABLE - a variable that does not have a fixed number of values. The variable REVENUE, for instance, can take on many different values.

CONTROL VARIABLE - the variable whose values are used to separate cases into subgroups. In a crosstabulation, this is the variable whose values form subtables.

CPU - Central Processing Unit, the locomotive of the computer that processes requests for computation. We care about the speed (expressed in Mhz) at which the CPU carries out these computations.

CROSSTABS - The command which orders SPSS/PC+ to carry out a crosstabulation.

CROSSTABULATION - the computation of the values of a variable as limited by the values of another variable: sector of business by location of business.

CURSOR - that blinking square or line on the screen that tells us where the keyboard input will appear on the screen.

DATA - information organized for analysis.

DEFAULT - the thing which the software will do if you don't tell it to do something else.

DISCRETE VARIABLE - a variable that has a limited number of values, which can have nominal or ordinal properties. REVENUE expressed in three values (High, Medium and Low) is a discrete variable.

DISC or DISK - a magnetic medium which stores data: a floppy, a hard disc drive, 5 1/4", 3 1/2".

DRIVE - the mechanism that spins a disc and enables the retrieval of specific bits of data stored on the disc.

EDITOR - any software program which allows you to enter text or data from the keyboard into the computer, to edit text or data, and to save files on disc. SPSS/PC+ contains an editor called REVIEW. WordPerfect is an editor. MS-DOS contains an editor.

ENTER - the large key, usually located on the right hand side of the middle line of the keyboard. We usually have to press the ENTER key to tell a computer to do something.

ESCAPE - the Esc key, usually located in the upper left hand corner of the keyboard; we press the Esc key to undo something, or to stop a process, or when we don't know what else to do. It usually offers an avenue of escape.

F10 - Function 10 key, usually located at the right end of the upper row of keys, or in the lower left hand corner. The Function keys usually have specific powers in computer programs. In WordPerfect, F7 tells the program to SAVE a file and EXIT. In SPSS/PC+, F10 has the power to tell the program to carry out a command, when followed by ENTER.

FASTGRAF - the name of a graphics package included in SPSS/PC+.

FILE - a container of specific data, which carries a unique name.

FILES, SPSS - to review these, SEE Chapter II, A - Welcome to SPSS

- Scratch Pad
- Listing File
- Log File
- System File
- Portable File
- Results File

FORMAT - as a verb, it means to prepare a disc to hold data. As a noun it means the way values of a variable are represented to the computer and the way these values are displayed. The DATA LIST and FORMATS commands assign formats that SPSS/PC+ uses to write or display values. The components of a format are the variable type (string or numeric), the variable width, and the number of decimal places.

FREQUENCIES - how many or how often data takes on a particular value.

GIGO - "Garbage In = Garbage Out", a computer operator's version of the conservation of energy, or, "you can't make a silk purse out of a sow's ear."

INCLUDE - a file of SPSS/PC+ commands and/or data to be processed in an SPSS/PC+ session via the INCLUDE command.

LABEL - a string that contains an extended description associated with a file, a variable, or a value of a variable.

LIS FILE - the same as a listing file: it contains the statistical and tabular output from SPSS/PC+ procedures, diagnostic information, and messages about the session. The default listing file is always SPSS.LIS. You can specify another file name using the SET command: SET /LISTING 'CAMEROON.LIS'.

LOG FILE - a file that contains all commands entered and processed by SPSS/PC+. The default log file is called SPSS.LOG, but you can call it something else, using the SET command.

MEMORY - the space available to hold data; in the world of computers, there is never enough memory space.

MISSING VALUE - you will see this in your output. This is a code that indicates that the true value of the variable could not be obtained, or a code that you would like to have ignored during statistical calculations.

NUMERIC VARIABLE - a variable whose values are numbers.

OUTPUT - the results produced: could be useful or nonsense, depending on your INPUT and how you manipulated it. In statistical analysis, as nowhere else, the GIGO rule is supreme.

OUTPUT FILE - the place to store the results produced

PLOT - a command which orders SPSS/PC+ to construct a graph out of your data

SESSION - a unit of time and labor that starts when you issue SPSS/PC+ the first command and ends with the command FINISH. F10 ENTER

SIMPLE VARIABLE - a variable that, in our case, relates directly to a single question in the survey.

STRING VARIABLE, or, ALPHANUMERIC VARIABLE - a variable whose values contain letters or special characters as well as numbers.

SUBCOMMAND - additional and more detailed instructions included on a command line to tell SPSS/PC+ precisely what you want it to do.

SYS FILE - the system file: a binary file (made up of lots of 0's and 1's) specially formatted by SPSS/PC+, containing both data and the dictionary that is written by the SPSS/PC+ SAVE command.

TARGET VARIABLE - a variable that contains the results of a transformation.

TRANSFORMATION - to change the values of a variable to correct coding errors, to modify the coding scheme, to create new variables, or to construct an index. Very useful, especially when you get new ideas long after the original data has been encoded.

VALUE - a numeric or alphanumeric code that represents the status of a case on a variable.

VARIABLE - an observable entity which can take on more than one value or characteristic. Human beings are variables, as are all things that change in value: not a CONSTANT.

VARIABLE NAME - any name of up to 8 characters in length (the first of which must be a letter!) assigned to a variable. We used, creatively enough, the names V1, V2, V3, etc. We could have used names representing the value itself: Age, Length, Color, Rock, etc etc.

VARIABLE DEFINITION - the portion of the DATA LIST command that assigns a name to each variable you intend to analyze and provides information about the location and format of the variables in the data file.

VDT - video display terminal - the screen

WINDOW - when the VDT is divided into smaller frames, those frames are known as windows.

WORKSPACE - the amount of memory available to SPSS/PC+ to do what you want it to do for you.

APPENDIX A: CAMEROON SURVEY QUESTIONNAIRE AND FREQUENCY DISTRIBUTIONS

(Percentages in Parentheses, N=408)

NOTE: percentages may not add up to 100 due to rounding error

CAMEROON SURVEY

The purpose of the following survey is to gather information about the private sector in Cameroon, the beliefs held by business persons on different aspects of doing business, and the general investment climate. This study is financed by the U.S. Agency for International Development (USAID). This information will assist USAID/Cameroon in formulating its development strategy for the coming years. The information obtained here will be treated in a private and confidential manner. Nevertheless, questions deemed inappropriate do not have to be answered.

- Position of the person interviewed: (1). Owner
(2). Manager
(3). Professional/Technical
(4). Other

Date the survey was performed: _____
(day/month/year)

Name of Interviewer: _____

Approved by: _____

Questionnaire No. _____

1. Sampling source (CIRCLE ONE):

- (1). Formal registry (75) (2). Informal sampling (25)

2. Location of the firm (CIRCLE ONE):

- (1). Douala (54) (2). Limbe (2) (3). Tiko (0.5) (4). Edea (0.2) (5). Yaounde (27)
(6). Bafoussam (5) (7). Garoua (3) (8). Bamenda (5) (9). Ngaoundere (3)

SECTION 1: GENERAL INFORMATION ON THE FIRM

3. How important were the following when selecting this site for your firm?
 (FOR EACH CATEGORY PUT EITHER 1=IMPORTANT; 2=NOT IMPORTANT; 3=DON'T KNOW)

_____		1	2	3
_____	a). Proximity to raw materials	(36)	(59)	(5)
_____	b). Proximity to labor	(32)	(63)	(5)
_____	c). Proximity to clients for my products	(72)	(26)	(2)
_____	d). Tax incentives	(14)	(67)	(18)
_____	e). Personal considerations	(54)	(39)	(6)
_____	f). Other (specify): _____	(13)	(28)	(59)

4. How long has the firm been operating? 10 years

5. What is the gender of the owners (or the majority shareholders) of this firm (CIRCLE ONE RESPONSE)

(1). Male (2). Female (3). Equal proportion of shares held by male and female (4). Don't Know

(85) (7) (4) (4)

6. How many permits, licenses did you have to obtain to establish (start up) your business? (CIRCLE ONE)

(1). 0-5 (2). 6-10 (3). 11-15 (4). 16-20 (5). Over 20

(56) (24) (10) (3) (5)

7. Approximately, how long did it take you to obtain all the paperwork necessary to begin business operations? (CIRCLE ONE)

(1). 0-6 months (2). 7-12 months (3). 13 to 18 months

(56) (24) (8)

(4). 19-24 months (5). Over 24 months

(4) (8)

8. Approximately, how many permits and licenses do you need to operate your business? (CIRCLE ONE)

(1). 0-5 (2). 6-10 (3). 11-15 (4). 16-20 (5). Over 20

(56) (24) (8) (4) (8)

9. How many of the permits and licenses you need to operate need to be renewed annually:
- (1). 0-5 (84) (2). 6-10 (21) (3). 11-15 (7) (4). 16-20 (2) (5). Over 20 (6)

10. In which sectors does your business operate? (Mark ALL categories either 1= YES or 2=NO):
- (6) a). Agriculture, forestry, fishing production (IF YES, GO TO NO. 11)
- (10) b). Processing of agricultural, forestry, fish products (IF YES GO TO NO. 12)
- (47) c). Manufacturing (IF YES GO TO NO. 13)
- (49) d). Trade and Commerce (IF YES GO TO NO. 14)
- (22) e). Other Services (IF YES GO TO NO. 15)
- (1) f). Mining (IF YES GO TO NO. 16)

11. In what MAIN agricultural activity is your business involved?: (CIRCLE ONE)

- | | |
|----------------------------------|--------------------------------|
| (01). Coffee (8) | (07). Livestock (20) |
| (02). Cocoa (4) | (08). Aquaculture/Fishing (16) |
| (03). Bananas/Plantains (0) | (09). Cotton (0) |
| (04). Other fruits (4) | (10). Palm Oil (0) |
| (05). Vegetables (4) | (11). Rubber (0) |
| (06). Rice (4) | (12). Tobacco (0) |
| | (13). Forestry/logging (36) |
| (14). Other (specify): (4) _____ | |

(GO TO NO. 16)

12. In what MAIN agricultural processing activity are you involved? (CIRCLE ONE)

- | | |
|---------------------------------------------------------|---------------------------------|
| (1). Cocoa/Coffee processing
(12) | (5). Ginning/Weaving
(2) |
| (2). Processed foodcrops
(fruits/vegetables)
(15) | (6). Palm oil processing
(2) |
| (3). Animal products
(meat and dairy)
(12) | (7). Tobacco processing
(2) |
| (4). Hides and skins
(0) | (8). Wood processing
(42) |
| (9). Other (specify): <u>12</u> | |

(Go to question no. 16)

13. In what MAIN manufacturing activity are you involved? (CIRCLE ONE):

- | | |
|--------------------------------------------------|-----------------------------------------------|
| (01). Handicraft
(4) | (08). Assembly (other)
(3) |
| (02). Clothing/textiles
(10) | (09). Brewing and beverage
products
(6) |
| (03). Finished wood products
(9) | (10). Machinery and equipment
(2) |
| (04). Plastics
(6) | (11). Paper and Paper products
(9) |
| (05). Metal products
(15) | (12). Leather/footwear
(3) |
| (06). Confectionery/baking
(17) | (13). Chemical/pharmaceutical
(7) |
| (07). Assembly (electrical
electronic)
(2) | (14). Other (specify): <u>(8)</u> |

(Go to question no. 16)

14. In what MAIN commercial activity are you involved? (Circle one):

- | | |
|------------------------------------|--------------------------------|
| (1). Wholesale Trade
(14) | (4). Direct import
(5) |
| (2). Retail Trade
(40) | (5). Export
(4) |
| (3). Both wholesale/retail
(34) | (6). Both import/export
(2) |
| (7). Other (specify): <u>(0.5)</u> | |

(Go to question no. 16)

15. In what MAIN service activity are you involved? (Circle one):

- | | |
|-------------------------------------------------|-----------------------------------------|
| (1). Vehicle and Appliances
Repairs
(17) | (5). Restaurant/food preparation
(4) |
| (2). Transportation
(13) | (6). Finance
(2) |
| (3). Communications
(1) | (7). Construction
(23) |
| (4). Tourism
(hotels/travel agencies)
(3) | (8). Consulting
(1) |
| (9). Other (specify): _____ (35) | |

(GO TO question no. 16)

16. What percentage of your direct sales (approximately) are to the following markets:

- | | |
|-------------------------------------------|-------------|
| a). Within the province | <u>71</u> % |
| b). Outside province, but within Cameroon | <u>21</u> % |
| c). within other UDEAC ⁽³⁾ | <u>3</u> % |
| d). To other parts of Africa | <u>0</u> % |
| e). To the EEC | <u>4</u> % |
| f). To other Europe | <u>1</u> % |
| g). To North America | <u>0</u> % |
| h). To Asia | <u>0</u> % |
| i). To the Middle East | <u>0</u> % |
| j). Other | <u>0</u> % |

³Includes Cameroon, Chad, CAR, Equatorial Guinea, Gabon, Congo.

17. Of your domestic sales, approximately what percentage of your product do you sell to:

a). Retail customers	<u>54</u> %
b). Small firms (less than 10 workers)	<u>13</u> %
c). Larger private firms	<u>18</u> %
d). Government agencies	<u>6</u> %
e). Parastatals	<u>3</u> %
g). Other	<u>6</u> %

18. What percentage of your product do you sell:

a). Through a cooperative:	<u>2</u> %
b). To an affiliated company or partner:	<u>6</u> %

19. What percentage of your raw materials (approximately) come from:

a). Within the province	<u>48</u> %
b). Outside province, but within Cameroon	<u>16</u> %
c). within Other UDEAC ⁽⁴⁾	<u>1</u> %
d). other parts of Africa	<u>2</u> %
e). EEC	<u>21</u> %
f). Other Europe	<u>3</u> %
g). North America	<u>0.5</u> %
h). Asia	<u>2</u> %
i). Middle East	<u>0.15</u> %
j). Other	<u>0.5</u> %

⁴Includes Cameroon, Chad, CAR, Equatorial Guinea, Gabon, Congo.

20. Approximately what percentage of the locally purchased raw materials are bought from:

- a). Parastatals 6 %
- b). Directly from farmers or individuals 14 %
- c). Small firms (less than 10 workers) 16 %
- d). Larger private firms 36 %
- e). Other 5 %

21. Is your firm owned by Cameroonians?(CIRCLE ONE)

- (1). YES (70) (If YES, no. 23)
- (2). NO (29) (if NO or Don't Know/NA, go to no. 22)
- (3). Don't know/not apply (1)

22. Approximately what percentage of the total equity⁵ is provided by NON-Cameroonians?

- a). French: 45 %
- b). British: 2 %
- c). German: 2 %
- d). US: 2 %
- e). Greek 7 %
- f). Lebanese 4 %
- g). Other Foreign: 25 %

23. Is part of the firm owned by Cameroonian public institutions (e.g. SNI or direct Government ownership)? (CIRCLE ONE)

- (1). YES (4) (if YES, no. 24)
- (2). NO (94) (If NO or Don't Know/NA, go to no. 25)
- (3). Don't know/not apply (1)

⁵Equity means non-loan money invested in the business.

29. What do you think will happen to your sales volume over the next year? (CIRCLE ONE)

(1). Will improve by MORE than 20 percent. (8)

(2). Will improve by LESS than 20 percent. (14)

(3). Will not change. (15)

(4). Will deteriorate by LESS than 20 percent. (10)

(5). Will deteriorate by MORE than 20 percent. (10)

(6). Don't know (43)

30. Please estimate the percentage of capacity at which your business:

a). Operates today: 49 % of capacity

b). Operated 3 years ago: 60 % of capacity

31. How have the following factors affected your sales volume over the past year?

(Mark ALL categories either 1=positive; 2=negative; 3=No effect; 4=Don't Know)

	1	2	3	4
___ a). Domestic demand for my product	(35)	(43)	(15)	(6)
___ b). Foreign demand for my product	(8)	(14)	(66)	(11)
___ c). Competition from other private firms in Cameroon	(10)	(52)	(31)	(7)
___ d). Competition from Cameroonian parastatals or government agencies	(3)	(17)	(67)	(13)
___ e). Competition from legally imported goods	(6)	(25)	(54)	(15)
___ f). Competition from illegally imported goods	(7)	(39)	(41)	(13)
___ g). Distance to markets	(9)	(14)	(66)	(11)
___ h). Market infrastructure (roads, port facilities, airport facilities)	(12)	(19)	(60)	(9)
___ i). Transportation facilities (availability of freight space)	(12)	(14)	(62)	(11)
___ j). Government regulations	(10)	(37)	(39)	(14)
___ k). Taxes	(7)	(47)	(36)	(10)
___ l). Exchange rates	(2)	(15)	(62)	(21)
___ m). Production technology	(17)	(14)	(54)	(14)

32. How did each of the government regulations/actions affect you business performance last year?

(Mark ALL categories either 1=positive; 2=negative; 3=No effect; 4=Don't Know)

	1	2	3	4
___ a). Permits, licenses requirements	(8)	(37)	(47)	(7)
___ b). Price/margin controls	(8)	(36)	(50)	(6)
___ c). Labor Code regulations on wages	(8)	(16)	(64)	(12)
___ d). Labor Code regulations of hiring/firing	(5)	(18)	(64)	(14)
___ e). Customs clearance procedures	(7)	(37)	(45)	(10)
___ f). Tax policies	(7)	(41)	(41)	(11)
___ g). The General Trade Schedule (policies on tariffs and quotas)	(6)	(23)	(51)	(18)
___ h). Investment Code incentives	(13)	(9)	(53)	(24)
___ i). Single Tax Regime (TU)	(10)	(14)	(48)	(28)
___ j). Internal Tax on Production Regime (TIP)	(7)	(15)	(50)	(27)

33. Which of the following fees, taxes, patents must your business pay?:

(Mark ALL categories either 1=yes; 2=no; 3=don't know; 4=NOT apply)

YES

- (66) a). Tax on internal turnover (TIT)
- (70) b). Registration fees for company charters (DESCA and DESBI)
- (77) c). Other registration fees and stamp duties regarding property leases (Patente, etc.)
- (36) d). Land, forest or mining taxes/licenses
- (65) e). Credit foncier deduction
- (42) f). Apprenticeship tax
- (40) g). Proportional tax on income from securities (TPRCM)
- (20) h). Taxes/duties on sale of securities/property
- (55) i). Direct/indirect taxes or duties on sales and purchases
- (45) j). Tax on insurance contracts
- (57) k). Trade union tax for employees
- (46) l). Fixed minimum corporate tax (IMFS)
- (38) m). Special corporate tax (TSS)
- (47) n). Corporate tax
- (56) o). Profit taxes
- (40) p). Community taxes
- (14) q). Other (specify): _____

SECTION 3: RESOURCE CONSTRAINTS

Land

34. What rights do you have to the land you are occupying?
(CIRCLE ONE)

- (1). Title deed (34) (2). Lease (62) (3). Squatter (2) (4). Other (2)

35. Please indicate how the following affected (or affect) your firm's operations over the last year: (Mark ALL categories either 1=positive; 2=negative; 3=No effect; 4=Don't Know)

- | | 1 | 2 | 3 | 4 |
|----------------------------------------------------------------|-----|------|------|------|
| ___ a). Cost of land | (5) | (22) | (66) | (7) |
| ___ b). Access to land | (2) | (13) | (76) | (10) |
| ___ c). Uncertainty about ability to stay on the land (tenure) | (5) | (23) | (63) | (9) |
| ___ d). Lack of clear titling | (3) | (12) | (72) | (13) |
| ___ e). Other (specify): _____ | (1) | (3) | (46) | (49) |

Labor

36. a). Number who work full time: 72

b). Number contracted as needed (on average): 14

37. How many of your (full time) employees are in each of the following categories?

	(1) Total Number	(2) Total Cameroonian	(3) Total Women
a). Managerial Personnel(⁶)	<u>3</u>	<u>2</u>	<u>0.3</u>
b). Other skilled Personnel(⁷)	<u>25</u>	<u>17</u>	<u>2</u>
c). Unskilled	<u>46</u>	<u>45</u>	<u>3</u>

⁶For example, managing directors, company executives, general managers.

⁷Includes professionals (for example engineers, accountants, economists), technical personnel (such as, mechanics, repairmen, plumbers, artisans, production line technicians, assemblers) and secretarial personnel.

38. What percentage of the top management team⁸ are women?
8 %
39. What percentage of the top management (the company decision-makers) in this firm is Cameroonian? 66 %
40. Are you satisfied with the productivity of your employees?
 (CIRCLE ONE)
- | | | |
|----------|---------|---------------------------|
| (1). YES | (2). NO | (3). Don't know/not apply |
| (79) | (9) | (12) |
41. On what basis do you pay your employees? (CIRCLE ONE)
- | | |
|------------------------------------------------------|------|
| (1). Based on the Labor Code statutory wage schedule | (57) |
| (2). Based on piece work or contractual basis | (20) |
| (3). Both | (14) |
| (4). Other | (9) |
42. Do you give bonuses or other incentives to get employees to work harder? (CIRCLE ONE)
- | | | |
|----------|---------|---------------------------|
| (1). YES | (2). NO | (3). Don't Know/not apply |
| (70) | (28) | (3) |

⁸Refers to the people in the firm involved in making the key strategic decisions about the firm's operations, NOT just the day to day production decisions.

43. For the following personnel categories, please specify how difficult it has been to obtain Cameroonians with the training/experience needed in your firm: (Mark ALL categories either 1=Difficult; 2=Average; 3=Easy; 4=Don't know/Not apply)

	1	2	3	4
___ a) Managerial Personnel ⁽⁹⁾	(21)	(14)	(39)	(27)
___ b) Secretarial/clerical Personnel	(6)	(11)	(56)	(27)
___ c) Professional Personnel ⁽¹⁰⁾	(17)	(16)	(36)	(31)
___ d) Technical Personnel ⁽¹¹⁾	(24)	(19)	(33)	(23)

44. Do you have to pay very high salary and benefits packages to attract: (Mark ALL categories either 1=yes; 2=no; 3=don't know/not apply)

YES

(38) a) good managers? ___(36) b) good professionals?

(42) c) good technical personnel?

(If answered YES to any of the above, go to no. 45, otherwise go to no. 46)

45. Has the need to pay high salary and benefits packages hindered the growth of your business? (CIRCLE ONE)

(1). YES (2). NO (3). Don't Know/Not apply
 (19) (69) (12)

Credit

46. Have difficulties obtaining access to credit affected your business performance over the last year? (CIRCLE ONE)

(1). YES (2). NO (3). Don't Know/not apply
 (57) (28) (15)

⁹For example, managing directors, company executives, general managers.

¹⁰For example: engineers (agricultural, chemical, etc.), accountants, economists.

¹¹For example, mechanics, repairmen, plumbers, artisans, production line technicians, assemblers.

47. Please specify how difficult you think it has been to obtain each type of loan over the last year? (Mark ALL categories either 1=Difficult; 2=Average; 3=Easy; 4=Don't know/Not apply)

___ a)	Short-term loans for working capital (e.g. less than 1 year)	(45)	(13)	(13)	(29)
___ b)	Medium-term loan for plant and equipment (e.g. 1-5 years)	(45)	(12)	(6)	(37)
___ c)	Long-term loan (e.g. over 5 years)	(47)	(6)	(5)	(42)

48. Please provide an estimate of the percentage of your capital (long and short term) which comes from the following sources:

a) Local commercial banks	<u>23</u> %
b) Other formal financial institutions	<u>3</u> %
c) Foreign sources	<u>10</u> %
d) Remittances	<u>0.7</u> %
e) Tontines	<u>12</u> %
f) Credit Unions	<u>2</u> %
g) Supplier's credit	<u>9</u> %
h) Family/friends	<u>7</u> %
i) Personal	<u>33</u> %

49. Please indicate how the following factors pertaining to bank policies affected your ability to access financing: (For each category of response, circle the appropriate number)

	1. Very Negative	2. Negative	3. Positive	4. Not at all	5 Don't Know/NA
a. Collateral requirements ⁽¹²⁾				1 (18)	2 (25)
b. Interest rates for short-term capital				3 (6)	4 (23)
c. Interest rates for long-term financing				5 (15)	(27)
d. Bank fees				(8)	(27)
e. Credit ceilings				(17)	(26)
f. Paperwork/Administrative procedures				(5)	(29)
g. Availability of venture capital sources				(5)	(17)
h. Distance to banks				(9)	(37)
				(2)	(4)
				(7)	(65)
				(22)	(22)
				(24)	(24)
				(30)	(30)
				(29)	(29)
				(33)	(33)
				(22)	(22)

Raw Materials

50. Have you had difficulty obtaining the raw materials that you need to operate? (CIRCLE ONE)

(1). YES (if YES go to no. 51)	(2). NO (if NO or Don't Know/NA, go to no. 52)	(3). Don't know/Not apply (if NO or Don't Know/NA, go to no. 52)
(35)	(56)	(9)

¹²Collateral means: security required by money lenders.

51. How have the following factors affected your firm's ability to procure all the inputs it needs: (For each category of response, circle the appropriate number)

1. Very Negative	2. Negative	3. Positive	4. Not at all	5 Don't Know/NA			
			1	2	3	4	5
a) Availability of financing	(34)	(37)	(8)	(15)	(6)		
b) Lack of cooperative buying	(8)	(15)	(5)	(47)	(25)		
c) Availability of transportation			(10)	(23)	(15)	(42)	(10)
d) Cost of transportation	(14)	(25)	(14)	(37)	(10)		
e) Import licensing controls	(9)	(20)	(6)	(41)	(23)		
f) Foreign exchange regulations	(3)	(17)	(4)	(48)	(28)		
g) Customs regulations	(8)	(26)	(6)	(40)	(19)		
h) Middlemen margins	(5)	(13)	(8)	(43)	(31)		
i) Scarcity/shortages of products			(18)	(32)	(8)	(32)	(10)
j) The price of the input	(11)	(32)	(4)	(34)	(19)		
k) Enforcement of supply contracts	(8)	(18)	(17)	(34)	(24)		

52. Is reliable information on the LOCAL market/demand for your product available? (CIRCLE ONE)

- (1). YES (55) (2). NO (35) (3). Don't know/Not apply (10)

53. Is reliable information on the INTERNATIONAL market/demand for your product available? (CIRCLE ONE)

- (1). YES (19) (2). NO (51) (3). Don't know/Not apply (31)

54. How interested are you on the following types of market information for your business? (For each category of response, circle the appropriate number)

	1. Interested	2. Not Interested	3. Not apply			4. Don't know
			1	2	3	
a) Information on prices			(89)	(6)	(2)	(2)
b) Information on product standards			(79)	(9)	(6)	(6)
c) Contacts with other companies			(80)	(7)	(7)	(5)
d) Information on trade regulations (tariffs/quotas)			(76)	(9)	(4)	(7)
e) Other (specify): _____			(12)	(4)	(8)	(76)

55. Specify the sources which you use to get information on the market for your product: (Mark ALL categories either 1=yes; 2=no)

- YES
- (36) a) The Ministry of Trade
 - (92) b) Personal Contacts
 - (29) c) Foreign Partners
 - (22) d) Syndustricam
 - (41) e) Chamber of Commerce
 - (10) f) CAPME
 - (59) g) Trade journals or newspapers
 - (5) h) National Marketing Board
 - (13) i) Other (specify): _____

Technology

56. Do you have access to appropriate production technology for your business?

- (1). YES (59)
- (2). NO (28)
- (3). Don't know/not apply (13)

(IF answered NO, go to No. 57, otherwise go to No. 58)

57. Have any of the following hindered access to needed technology (Mark ALL categories either 1=yes; 2=no; 3=don't know/not apply)

YES

(51) a) Lack of information about available technology

(37) b) Lack of local expertise/experience

(49) c) Cost of technology

(26) d) Reluctant to invest large sums in fixed assets

(18) e) Imported technology not adequate for the market

(19) f) Government licensing requirements

(35) g) Difficulty access to spare parts

(42) h) Maintenance problems

(2) i) Other (specify): _____

Infrastructure

58. How have each of the following affected your ability to produce and market your product?: (Mark ALL categories either 1=negative; 2=positive; 3=Don't Know/not apply)

	1	2	3	4
___ a) Availability of air cargo space	(5)	(8)	(78)	(10)
___ b) Price of air cargo space	(10)	(8)	(73)	(9)
___ c) Availability of land transport facilities	(19)	(25)	(48)	(7)
___ d) Price of land transport facilities	(21)	(26)	(47)	(6)
___ e) Availability of maritime transport facilities	(17)	(10)	(64)	(9)
___ f) Price of maritime transport	(15)	(9)	(65)	(10)
___ g) Existence of an adequate road network	(27)	(22)	(44)	(5)
___ h) Quality of road network	(29)	(21)	(43)	(7)
___ i) Transport regulations (on weight, axles, etc.)	(12)	(14)	(61)	(13)
___ j) Road checks/road blocks	(33)	(13)	(47)	(7)
___ k) Cost of water	(15)	(17)	(61)	(7)
___ l) Access to/reliability of water supply	(12)	(14)	(66)	(8)
___ m) Cost of electricity	(46)	(22)	(27)	(5)
___ n) Access to/reliability of electricity	(27)	(19)	(49)	(6)
___ o) Cost of telecommunications (telephone, fax, etc)	(36)	(17)	(40)	(7)
___ p) Access to/reliability of telecommunications	(31)	(15)	(46)	(8)

Legal Environment

59. Given your experience as a business person, please tell us if you agree with the following statements regarding your experience with the legal system: (Mark ALL categories either 1=Agree; 2=Disagree; 3=Don't Know/Not apply)

Agree

- (61) a) Contracts are difficult to enforce.
- (63) b) Problems enforcing contracts make it difficult to do business in Cameroon.

SECTION 4: OPPORTUNITIES

60. Have you formulated a specific plan to invest in any of the following areas? (Mark ALL categories either 1=yes; 2=no; 3=don't know/not apply)

- YES
- (50) a) Personnel training and development
 - (57) b) Improved production technology
 - (52) c) physical capital (plant improvement)
 - (71) d) Improving my own management capability
 - (73) e) Marketing
 - (65) f) Procurement
 - (71) g) Quality Control

Suppose you were considering investing in one of the areas/activities listed below:
(Read or show the list)

	WORST	BEST
(01). Unprocessed agriculture production for domestic market	(15)	(5)
(02). Unprocessed agriculture production for export market	(23)	(4)
(03). Agroprocessing/agroindustry for domestic market	(4)	(13)
(04). Agroprocessing/agroindustry for export	(5)	(9)
(05). Manufacturing for export	(11)	(13)
(06). Manufacturing for domestic market	(7)	(13)
(07). Construction	(15)	(4)
(08). Tourism	(10)	(8)
(09). Commerce and trade	(3)	(21)
(10). Banking	(6)	(8)
(11). Other services	(2)	(1)
(12). Other (specify): _____		(2)

61. Which one do you think offers the LOWEST or WORST return on investment?: _____ (put the number corresponding to the area listed above)

62. Which one do you believe offers the HIGHEST or BEST return on investment?: _____ (put the number corresponding to the area listed above)

(If answered 1, 2, 3, 4, go to No. 63, otherwise go to No. 64)

63. Which agricultural products do you think offer the BEST return on investment? (CIRCLE ONE)

- | | |
|---------------------------------|-------------------------------------|
| (01). Coffee
(5) | (08). Aquaculture/Fishing
(4) |
| (02). Cocoa
(2) | (09). Cotton
(1) |
| (03). Bananas/Plantains
(14) | (10). Palm Oil
(10) |
| (04). Other fruits
(9) | (11). Rubber
(7) |
| (05). Vegetables
(3) | (12). Tobacco
(3) |
| (06). Rice
(3) | (13). Forestry/logging
(19) |
| (07). Livestock
(17) | (14). Other (specify): _____
(2) |

64. Are you currently exporting or have plans to export in the future? (CIRCLE ONE)

- | | | |
|------------------|-----------------|----------------------|
| (1). YES
(40) | (2). NO
(50) | (3). Not Sure
(9) |
|------------------|-----------------|----------------------|

(If answered YES, go to No. 65, otherwise go to No. 72)

(NOTE: if you are interviewing informal sector firms, skip now to question no. 74)

65. Which export market do you believe would be most profitable? (Circle ONE)

- | | |
|-----------------------|---------------------------------|
| (1). UDEAC (58) | (5). USA (5) |
| (2). Other Africa (7) | (6). Asia (1) |
| (3). EEC (19) | (7). Middle East (3) |
| (4). Other Europe (4) | (8). Don't Know/No Response (4) |

66. How important are the following as obstacles to entering/expanding your export operations in this market: (For each category of response, circle the appropriate number)

	1. Very Important	2. Somewhat Important	3. Not Important		4. Don't Know	
			1	2	3	4
a) Foreign exchange controls			(34)	(20)	(37)	(9)
b) Credit controls			(40)	(24)	(20)	(16)
c) Bureaucratic delays			(47)	(19)	(18)	(17)
d) Taxes			(34)	(29)	(26)	(11)
e) Production technology			(35)	(18)	(30)	(17)
f) Transport infrastructure			(48)	(21)	(21)	(10)
g) Other market infrastructure			(24)	(22)	(30)	(23)
h) Packaging			(26)	(24)	(36)	(14)
i) Raw materials			(39)	(13)	(28)	(20)
j) Labor productivity			(41)	(28)	(18)	(13)
k) Market information			(60)	(24)	(7)	(9)
l) Other (specify): _____			(7)	(5)	(0)	(87)

67. Have you heard about the Industrial Free Zone (IFZ) program being initiated in Cameroon? (CIRCLE ONE)

(1). YES (96) (2). NO (4)

68. In order to qualify for the incentives offered under the IFZ regime, a firm must be 100 percent export oriented. Would consider establishing an export oriented operation under the IFZ regime? (CIRCLE ONE)

(1). I already export 100 percent, and would be interested in learning more about qualifying for the IFZ incentives. (7)

(2). I already export 100 percent, but I am not interested in the IFZ regime. (6)

(3). I do not export 100 percent, but would be interested in establishing an export oriented operation under the IFZ regime. (55)

(4). I do not export 100 percent and I am NOT interested in the IFZ regime. (31)

(If answered 1 or 3, go to question no. 69, otherwise go to question no. 72)

69. The IFZ regime is designed to allow for development of an industrial park, as well as the establishment of single factory free zones, called "point francs." Would your operation require "point franc status" or could you locate your operation in an industrial park? (CIRCLE ONE)

(1). Require "point franc" status (49)

(2). Willing to locate in an industrial park. (25)

(3). Don't Know (27)

(If answered 2, go to NO. 70, otherwise move to question 72)

70. Would you prefer to rent or own a building? (CIRCLE ONE)

(1). Rent (17) (2). Own (61) (3). Don't know (22)

71. How much space would your business require if you were to locate in an IFZ? 17,984 sq. meters

72. Are you interested in joint venture with foreign companies, either for local production or export? (CIRCLE ONE)

(1). YES (56) (2). NO (35) (3). Don't Know (10)

(if YES, go to no. 73)

(if NO or Don't know, go to no. 74)

73. What resources do you expect from a foreign partner? (Mark ALL categories either 1=yes; 2=no; 3=don't know)

- YES
(87) a. Capital
(84) b. Technological expertise
(46) c. Managerial expertise
(90) d. Access to markets
(8) e. Other (specify): _____

SECTION 5: ASSOCIATIONS

74. Are you (or is your firm) affiliated with any type of business or trade association? (CIRCLE ONE)

- (1). YES (2). NO (3). Unsure
(42) (56) (2)
(if YES, (if NO or Unsure, no. 75)
no. 76)

75. Specify which one BEST describes why you do not belong to an association: (CIRCLE ONE)

- (1). No association provides services I find useful.
(56)
(2). I am unwilling to pay membership dues.
(2)
(3). I am not able to pay membership dues.
(15)
(4). Membership restrictions
(9)
(5). Other restrictions
(3)
(6). Other (16)

(Skip to question no. 77)

76. Please specify which associations and rate how effectively they have represented your business interests: (For each category of response, circle the appropriate number)

	1. Very Effective	2. Somewhat Effective	3. Not Effective	4. Don't Belong		
			1	2	3	4
a. Syndustricam			(24)	(20)	(5)	(51)
b. Chamber of Agriculture			(2)	(6)	(10)	(82)
c. Chamber of Commerce			(21)	(35)	(11)	(33)
d. Agriculture Commodity Associations			(2)	(2)	(6)	(90)
e. Hotel/Restaurateur Association			(2)	(0)	(5)	(93)
f. Groupement des Hommes D'Affaires Camerounais			(10)	(13)	(5)	(72)
g. Cooperative			(3)	(2)	(7)	(88)
h. Other (specify): _____			(14)	(10)	(4)	(72)

77. Which of the following services would you want a business/trade association to offer (or increase)? (For each category of response, circle the appropriate number)

	1. Very Interested	2. Somewhat Interested	3. Not Interested	4. Don't Know		
			1	2	3	4
a. Provide access to credit			(71)	(12)	(9)	(8)
b. Provide technical assistance			(48)	(21)	(21)	(9)
c. Develop personnel training services			(44)	(23)	(20)	(13)
d. Provide feasibility studies			(35)	(22)	(25)	(17)
e. Increase contact/representation with Government			(49)	(22)	(15)	(14)
f. Provide information on government administrative procedures/regulations			(49)	(25)	(13)	(15)
g. Provide information on export markets			(44)	(21)	(22)	(14)
h. Provide information on local market			(68)	(20)	(5)	(6)

THANK YOU FOR YOUR COOPERATION WITH THIS STUDY

APPENDIX B - List of Variables used in the Cameroon Survey

Part A - list of basic variables contained in the survey.

Part B - list of various basic data transformations.

Part C - list of the basic frequency distributions, the means and the cross tabs expected.

In part B there are a number of new variables in Part B (e.g V81, V94, V96) which are categorizations of continuous variables. For example, V81 is a new variable (to be labeled "firm size"), created by categorizing a continuous variable (V80 - no. of full time employees). Though for V81 as well as V94 and V96, I have suggested categories which I think make sense, given what we have seen in the sampling frame, you may want to change the various cutoff rules after you see what the actual distribution of the responses look like, since we want to create categories which are more or less evenly divided and also have some substantive meaning. Thus the categories may change, depending on the final distribution of these variables. You could try separating the firms out in the categories I suggest below and running some cross tabs I specified. If the results are significant (meaning that these categories yield significant differences in response patterns), then we are on the right track, otherwise we may have to re-categorize them.

Part A

Simple Variables (from the questionnaire)

1. V1 - Sampling source
2. V2 - Location
3. V3a - Proximity raw materials
V3b - Proximity labor
V3c - Proximity market
V3d - Tax incentives
V3e - Other location
4. V4 - Years operating
5. V5 - Gender owner
6. V6 - Months before starting
7. V7 - Regulations for starting
8. V8 - Regulations for operating
9. V9 - Annual renewals
10. V10a - Agriculture
V10b - Agroprocessing
V10c - Manufacturing
V10d - Commerce
V10e - Other services
V10f - Mining
11. V11 - Ag activity
12. V12 - Ag-process activity
13. V13 - Manuf activity
14. V14 - Commerce activity
15. V15 - Service activity

- 16. V16a - Sales province
V16b - Sales Cameroon
V16c - Sales UDEAC
V16d - Sales other Africa
V16e - Sales EEC
V16f - Sales other Europe
V16g - Sales America
V16h - Sales Asia
V16i - Sales MidEast
V16j - Sales other
- 17. V17a - Dom retail
V17b - Dom small firms
V17c - Dom large firms
V17d - Dom parastatals
V17e - Dom other
- 18. V18a - Sales coop
V18b - Sales affiliated
- 19. V19a - Inputs province
V19b - Inputs Cameroon
V19c - Inputs UDEAC
V19d - Inputs other Africa
V19e - Inputs EEC
V19f - Inputs other Europe
V19g - Inputs America
V19h - Inputs Asia
V19i - Inputs MidEast
V19j - Inputs other
- 20. V20a - Inputs parastatals
V20b - Inputs farmers
V20c - Inputs small private
V20d - Inputs large private
V20e - Inputs other
- 21. V21 - Nationality owner
- 22. V22a - French owner
V22b - UK owner
V22c - FRG owner
V22d - US owner
V22e - Greek owner
V22f - Lebanese owner
V22g - Other owner
- 23. V23 - Public owner
- 24. V24 - Percentage public
- 25. V25 - Owner structure

- 26. V26 - Business climate
- 27. V27 - CA
- 28. V28 - Past sales
- 29. V29 - Future sales
- 30. V30a - Capacity past
V30b - capacity present
- 31. V31a - Dom demand
V31b - foreign demand
V31c - local private
V31d - parastatals
V31e - legal imports
V31f - illegal imports
V31g - Distance market
V31h - infrastructure
V31i - transport
V31j - regulations
V31k - taxes
V31l - exchange
V31m - technology
- 32. V32a - permits
V32b - price controls
V32c - wage regs
V32d - employment regs
V32e - customs
V32f - tax policy
V32g - tariffs
V32h - IC
V32i - TU
V32j - TIP

- 33. V33a - TIT
- V33b - DESCA
- V33c - Patente
- V33d - land taxes
- V33e - foncier
- V33f - apprenticeship
- V33g - TPRCM
- V33h - Property taxes
- V33i - sale taxes
- V33j - insurance
- V33k - trade union
- V33l - IMFS
- V33m - TSS
- V33n - corporate tax
- V33o - Profit tax
- V33p - community
- V33q - other

- 34. V34 - land rights

- 35. V35a - land cost
- V35b - land access
- V35c - tenure
- V35d - titling
- V35e - other

- 36. V36a - No. full time
- V36b - No. part time

- 37. V37a1 - total managerial
- V37b1 - total other skilled
- V37c1 - total unskilled
- V37a2 - Cameroonian managerial
- V37b2 - Cameroonian other skilled
- V37c2 - Cameroonian unskilled
- V37a3 - Women managerial
- V37b3 - Women other skilled
- V37c3 - Women unskilled

- 38. V38 - top managers women

- 39. V39 - top managers cameroonian

- 40. V40 - labor productivity

- 41. V41 - payment mode

- 42. V42 - incentives

- 43. V43a - access managers
V43b - access clerical
V43c - access professional
V43d - access technical
- 44. V44a - salary managers
V44b - salary professionals
V44c - salary technical
- 45. V45 - Effect salary
- 46. V46 - access to credit
- 47. V47a - Access short-term K
V47b - Access medium-term K
V47c - Access long-term K
- 48. V48a - commercial bank credit
V48b - other FI credit
V48c - foreign credit
V48d - remittances
V48e - tontines
V48f - credit unions
V48g - suppliers credit
V48h - family
V48i - personal
- 49. V49a - effect collateral
V49b - effect s-term interest
V49c - effect l-term interest
V49d - effect bank fees
V49e - effect credit ceilings
V49f - effect paperwork
V49g - effect venture capital
V49h - effect distance
- 50. V50 - Access to raw materials
- 51. V51a - availability financing
V51b - coop buying
V51c - transport access
V51d - transport cost
V51e - import licensing
V51f - for.exchange
V51g - customs
V51h - middlemen
V51i - shortages
V51j - input price
V51k - contract enforcement
- 52. V52 - local market info

- 53. V53 - export market info
- 54. V54a - price info
V54b - prod standards
V54c - contacts
V54d - trade regs
V54e - other info
- 55. V55a - Min of Trade
V55b - personal contacts
V55c - foreign partners
V55d - Syndustricam
V55e - Chamber of Commerce
V55f - CAPME
V55g - printed media
V55h - marketing board
V55i - other source
- 56. V56 - access to technology
- 57. V57a - lack of info tech
V57b - local expertise tech
V57c - cost tech
V57d - large investment tech
V57e - inappropriate tech
V57f - licensing tech
V57g - spare parts tech
V57h - maintenance tech
V57i - other tech
- 58. V58a - air access
V58b - air price
V58c - land access
V58d - land price
V58e - sea access
V58f - sea price
V58g - road net
V58h - road quality
V58i - transport regs
V58j - road checks
V58k - water price
V58l - water access
V58m - electricity price
V58n - electricity access
V58o - telecom price
V58p - telecom access
- 59. V59a - contracts difficult
V59b - legal effects

- 60. V60a - invest personnel training
- V60b - invest tech
- V60c - invest plant
- V60d - invest management
- V60e - invest marketing
- V60f - invest procurement
- V60g - invest quality

- 61. V61 - worst return

- 62. V62 - best return

- 63. V63 - best ag return

- 64. V64 - plans to export

- 65. V65 - best export market

- 66. V66a - export forex
- V66b - export credit
- V66c - export bureauc
- V66d - export taxes
- V66e - export tech
- V66f - export trans
- V66g - export infra
- V66h - export pack
- V66i - export raw mate
- V66j - export labor
- V66k - export info
- V66l - export other

- 67. V67 - know IFZ

- 68. V68 - interest in IFZ

- 69. V69 - point franc

- 70. V70 - leasing IFZ

- 71. V71 - space IFZ

- 72. V72 - joint venture

- 73. V73a - jv capital
- V73b - jv tech
- V73c - jv manage
- V73d - jv markets
- V73e - jv other

- 74. V74 - membership association

- 75. V75 - reason not member

- 76. V76a - Syndustricam rating
- V76b - Chamber of Ag rating
- V76c - Chamber of Commerce rating
- V76d - Ag commodity assoc. rating
- V76e - Hotel assoc rating
- V76f - Groupement rating
- V76g - Coop rating
- V76h - other group rating
- 77. V77a - access credit
- V77b - access tech assist.
- V77c - personnel training
- V77d - feasibility
- V77e - lobby
- V77f - info on regs
- V77g - info on exports
- V77h - info on local market

Part B

New Variables

- 78. New variable V80 (NO. FULL TIME EMPLOYEES) in the following manner

COMPUTE V80=(V36a*1.0) + (V36b*0.5)

- 79. New variable V81 (FIRM SIZE)⁽¹³⁾

IF (V80 LT 2) V81=1 (OWNER)
 IF (V80 GT 2 AND LTE 5) V81=2 (MICRO)
 IF (V80 GT 5 AND LTE 10) V81=3 (SMALL)
 IF (V80 GT 10 AND LTE 100) V81=4 (MEDIUM)
 IF (V80 GT 100) V81=5 (LARGE)

- 80. New variable V82 (PERCENTAGE MANAGEMENT):

COMPUTE V82=V37a1/V36a ("/" means "divided by")

- 81. New variable V83 (PERCENTAGE SKILLED):

COMPUTE V83=V37b1/V36a

- 82. New variable V84 (PERCENTAGE UNSKILLED):

COMPUTE V84=V37c1/V36a

¹³⁾ Note: LT means "less than"; LTE means "less than or equal to"; GT means "greater than"; GTE means "greater than or equal to."

83. New variable v85 (PERCENTAGE CAM MANAGEMENT):
COMPUTE v85=v37a2/v37a1
84. New variable v86 (PERCENTAGE CAM SKILLED):
COMPUTE v86=v37b2/v37b1
85. New variable v87 (PERCENTAGE CAM UNSKILLED):
COMPUTE v87=v37c2/v37c1
86. New variable v88 (PERCENTAGE WOMEN MANAGEMENT):
COMPUTE v88=v37a3/v37a1
87. New variable v89 (PERCENTAGE WOMEN SKILLED):
COMPUTE v89=v37b3/v37b1
88. New variable v90 (PERCENTAGE WOMEN UNSKILLED):
COMPUTE v90=v37c3/v37c1

89. New variable V91 (PERCENTAGE WKFORCE CAM):
COMPUTE V91=(V37a2+V37b2+V37c2)/V36a
90. New variable V92 (PERCENTAGE WKFORCE WOMEN):
COMPUTE V92=(V37a3+V37b3+V37c3)/V36a
91. New variable V93 (REG BURDEN)⁽¹⁴⁾
COMPUTE V93 (@SUM V33a to V33q)
92. New variable V94 (MARKET ORIENTATION)
IF (V16a LTE 050 AND V16b LTE 050) V94=1 (EXPORTERS)
IF (V16a GTE 050 OR V16b GTE 050) V94=2 (DOMESTIC)
93. New variable V95 (SECTOR)
IF (V10a EQ 1) V95=1 (AG)
IF (V10b EQ 1) V95=2 (AGPRO)
IF (V10c EQ 1) V95=3 (MANUF)
IF (V10d EQ 1) V95=4 (COM)
IF (V10e EQ 1) V95=5 (SERV)
IF (V10f EQ 1) V95=6 (MINE)
94. New variable V96 (Subsidiary)
IF (18b GT 50) V96=1 (SUBSIDIARY)
IF (18b LTE 50) V96=2 (NONSUBSIDIARY)

¹⁴) @SUM means "sum all values of all these variables)