UNIVERSITY OF EDINBURGH
Edinburgh Regional Computing Centre

User Note 25

(August 1986)

Title:

SPSSX on EMAS 2900 and EMAS-3

Author: Helen Talbot

Contact: Advisory service

Software Support Category: See Note 15

Synopsis

SPSSX, the Statistical Package for the Social Sciences, contains statistical procedures and data management facilities of particular use for survey analysis. This Note describes how to access it on the BUSH, EMAS and EMAS-A services.

Contents

1.	Statistical techniques	2
2.	Documentation	2
3.	Courses	2
4.	Advice	3
5.	The SPSSX command on EMAS 2900	3
6.	Accessing files with SPSSX on EMAS 2900	3
7.	The SPSSX command on EMAS-3	4
8.	Accessing files with SPSSX on EMAS-3	5
9.	Interactive and batch jobs on EMAS 2900	5
10.	Interactive and batch jobs on EMAS-3	8
11.	Errors on EMAS 2900 and EMAS-3	9
11.1	SPSSX Errors	6
11.2	EMAS 2900 Errors	6
12.	Examples of running SPSSX jobs on EMAS 2900	7
13.	Examples of running SPSSX jobs on EMAS-3	8
14.	Hints on running SPSSX more efficiently	9

Edinburgh Regional Computing Centre

James Clerk Maxwell Building, The King's Buildings, Mayfield Road, Edinburgh, EH9 3JZ. Telephone 031-667 1081

1. Statistical techniques

The statistical techniques available are:

Analysis of variance and covariance
Bivariate correlation
Box Jenkins
Canonical correlation
Contingency tables
Discriminant analysis
Factor analysis
Frequency distributions
Life table analysis
Log-linear analysis
Multiple regression
Multiple response analysis
Non-parametric correlation
Non-parametric tests
Partial correlation
Reliability analysis
T-tests

In addition SPSS-X version 2 (available on EMAS-A only) includes:

Cluster analysis
Multidimensional scaling analysis
Probit analysis

2. Documentation

The SPSSX User's Guide (2nd edition, 1986) may be obtained from bookshops, or from CAST (below) or direct from:

SPSSX Benelux,
Avelingen West 5,
P.O. Box 115,
4200AC
Gorinchem,
Netherlands.

The SPSSX Introductory Guide, written by the Program Library Unit who maintain SPSSX on EMAS 2900, is available at a price of £2 from:

Miss J. Anderson,
CAST,
1 Roxburgh Street,
Edinburgh.
EH8 9TA

The SPSSX manuals can be consulted in the ERCC Advisory rooms at JCMB and 59 George Square, and also in the user terminal room in Appleton Tower basement. The EMAS command `PACKHELP SPSSX` gives information about running SPSSX jobs.

3. Courses

The Centre for Applications Software and Technology (CAST) offers courses on SPSSX each term to staff and post-graduate students. Contact the ERCC receptionist at 59 George Square for details, telephone 031-667 1011 ext. 2300.

4. Advice

The ERCC Advisory Service gives advice to users of SPSSX on EMAS. The Advisory Service is normally open as follows:

Room 3205 JCMB	Monday-Friday	9.30-12.30, 14.00-17.00	667 1081 ext. 2976/7
59 George Square	Monday-Friday	9.30-12.30, 14.00-17.00	667 1011 ext. 2300

The EMAS command `HELP ADVISORY` gives details of any variations to this schedule.

5. The SPSSX command on EMAS 2900

To obtain access to SPSSX (and other packages mounted by the Program Library Unit) type the following command:

```
OPTION SEARCHDIR=PLULIB.PACKDIR
```

Unless the file SS#OPT has been archived because you have not been using your process for several weeks, it should not be necessary to type this command again.

The SPSSX command is of the form:

```
SPSSX PARAMETER1=REPLY1,PARAMETER2=REPLY2,...
```

The parameters can be in any order and can be abbreviated to the first three letters. The possible parameters and replies are listed in the following table.

Table 1

Input Parameters for SPSSX on EMAS 2900

Parameter Name	Default	Reply
CONTROL	.IN	The name of the file containing the SPSSX instructions.
LISTING	.OUT	Filename or device for output. For a line printer this should be of the form .LPnn where nn is the printer number. e.g. .LP15 is the JCMB printer, .LP25 is the Appleton Tower printer.
WORKSPACE	256	The workspace size in kilobytes. The default is usually adequate and so this parameter is not usually needed
MAXWORK	256	Maximum workfile size in Kbyte. A guide to the value is 8 times number of cases times number of variables. The default is usually adequate and so this parameter is not usually needed.

6. Accessing files with SPSSX on EMAS 2900

Provided a filename conforms to the following format, the accessing of files in SPSSX is very simple:

1. The name must not have more than 8 characters.
2. The first character in the name must be a letter.
3. All other characters must be either letters or numbers.
4. The file must be in the user's process.
5. It must not be partitioned.

If the file name conforms to this then to read an SPSS or SPSSX system file called FRED use:

```
GET FILE=FRED
```

To save data in an SPSSX system file use:

```
SAVE OUTFILE=FRED
```

To read data in from an ordinary data file use the SPSSX DATALIST command. Thus:

```
DATALIST FILE=FRED
```

If the file does not conform to the naming conventions above or belongs to another user then an SPSSX FILE HANDLE command must be used. e.g.

```
FILE HANDLE FRED/NAME='EXPL55.FREDDY'
GET FILE=FRED
```

will get an SPSSX system file called FREDDY in the process of user EXPL55. A FILE HANDLE statement may be used even when unnecessary according to the above rules. Some versions of SPSSX always require a FILE HANDLE statement.

In the case of output, files must not be members of partitioned files and it is not possible to create or extend another user's files. Magnetic tape files cannot be used as output devices by the present version of SPSSX.

7. The SPSSX Command on EMAS-3

To obtain access to SPSSX (and other packages mounted by the Program Library Unit) type the following command:

```
SEARCHDIR PLULIB:PACKDIR
```

This command need only be issued once.

The SPSSX command is of the form:

```
SPSSX PARAMETER1=REPLY1,PARAMETER2=REPLY2,...
```

The parameters can be in any order and can be abbreviated to the first three letters. The possible parameters and replies are listed in the following table.

Table 2

Input Parameters for SPSSX on EMAS-3

Parameter Name	Default	Reply
CONTROL	.IN	The name of the file containing the SPSSX instructions.
LISTING	.OUT	Filename or device for output. For a line printer this should be of the form .LPnn where nn is the printer number. e.g. .LP15 is the JCMB printer, .LP25 is the Appleton Tower printer.
FH1		File Handle/EMAS filename } for files to be
FH2		File Handle/EMAS filename } accessed from SPSSX
FH3		File Handle/EMAS filename } (see Section 8).
DATA		File Handle/Raw data input file

8. Accessing files with SPSSX on EMAS-3

To access an SPSSX system file on EMAS-3 you must use one of the FH1, FH2 or FH3 parameters when you call SPSSX. Thus to access a system file called FREDDY which is not in your own process you would call SPSSX as:

```
SPSSX CONTROL=CONT, FH1=FRED/USERNN:FREDDY
```

USERNN is the user number
and the control file, CONT would contain the statement

```
GET FILE=FRED
```

To access a raw data file called JIMMY using a DATA LIST statement then you would call SPSSX as

```
SPSSX CONTROL=CONT, DATA=JIM/JIMMY
```

and the control file, CONT would contain a statement of the form

```
DATA LIST FILE=JIM
```

9. Interactive and batch jobs on EMAS 2900

SPSSX jobs can be run interactively by typing SPSSX at command level or by using OBEY or OBEYJOB with a file containing the SPSSX command. It is more practical with most jobs to run a batch job by using DETACH to put a job in the batch queue.

First edit a file to contain the SPSSX command. Then DETACH that file.
The DETACH command takes the form:

```
DETACH filename,cputime,paramfile
```

Only the filename is necessary.

If you expect the job to take more than the default time of 30 seconds, put the number of CPU seconds as the second parameter.

If you want to use any of the document parameters, which is usually the case, you

will need to either put a filename as the third parameter or put .IN and enter the document parameters in response to prompts, ending with .END. If the job is detached successfully, you will get a message at the terminal. This will give you the document number. Typing the command DOCUMENTS gives you a list of any outstanding batch jobs or files to be output. If you type DOCUMENTS ddd where ddd is the document number, you will get more information about the job. If you realize after you DETACHed a job that you made a mistake and the job is still in the queue, you can remove the job from the queue by typing the command DELETEDOC ddd where ddd is the document number.

One possible document parameter you might want to use is OUT, if you want to reroute the job journal to a file or a nearby printer.
The default for OUT is LP.
The information about a job running is contained in the job journal. To avoid time waiting for the output to be delivered to you if you know there has been an error (result code 8 instead of 0), use the OUT or OUTNAME parameter.

e.g. for a printer put
```
OUT=LPnn
```

e.g. for a file put
```
OUTNAME=filename
```

where `filename` denotes a file which must not already exist.

For a description of these document parameters and others, see Chapter 2 of the EMAS 2900 User's Guide.

10. Interactive and batch jobs on EMAS-3

SPSSX jobs can be run interactively by typing SPSSX at command level or by using OBEY or OBEYJOB with a file containing the SPSSX command. It is more practical with most jobs to run a batch job using DETACH or to put a job in the batch queue.

First edit a file to contain the SPSSX command: then DETACH that file.
The DETACH command takes the form:

```
DETACH filename,cputime,log
```

Only the `filename` is necessary.

If you expect the job to take more than the default time of 30 seconds, put the number of CPU seconds as the second parameter.

If the job is detached successfully, you will get a message at the terminal. This will give you the document number. Typing the command BATCH gives you a list of any outstanding batch jobs. If you type BATCH ddd where ddd is the document number, you will get more information about the job. If you realize after you DETACHed a job that you made a mistake and the job is still in the queue, you can remove the job from the queue by typing the command DELETEDOC ddd where ddd is the document number.

The information about a job running is contained in the job journal. By default this is directed to a line printer. if you want to reroute the job journal to a file or a nearby printer, then use the third parameter. To avoid time waiting for the output to be delivered to you if you know there has been an error (result code 8 instead of 0), use the `log` parameter.

e.g. for a printer put
```
detach spssjob1,,LPnn
```

e.g. for a file put
```
detach spssjob2,,filename
```

where `filename` denotes a file, which will be overwritten if it already exists.

For a description of these document parameters and others, see Chapter 2 of the EMAS 2900 User's Guide.

11. Errors on EMAS 2900 and EMAS-3

11.1 SPSSX errors

If the job has failed, you will get a message saying the result code is 8; 0 indicates a successful run.

Errors in SPSSX statements are marked with an error number where they occur in the output, together with an explanation in English.

11.2 EMAS 2900 and EMAS-3 errors

The following EMAS errors may occur during SPSSX runs.

(a) CPU TIME EXCEEDED
This occurs if the job takes longer than the maximum CPU time requested. The default for batch jobs is 30 seconds.
Action: DETACH the job again with a larger time limit as the second parameter.

(b) FILE NOT AVAILABLE
This occurs if the file you have asked for is not on your filespace. This may mean you have mistyped the name or that it has been archived or that it belongs to someone else who has not PERMITted you to access it.

(c) NO FREE FILE DESCRIPTORS
This usually occurs if there is not enough room for more file names in your file index but can also occur if you have run out of section or permission descriptors.
The command `FILES ,P` tells you the size of your file index and how many file, section and permission descriptors are left. One file descriptor is required for each file. One section descriptor is required for each 128K section of a file after the first. One permission descriptor is required for each permission granted to a user or group or users.
Action: DESTROY and/or ARCHIVE files to make more room in your index. You can also group some files into a partitioned file as partitioned files use only one file descriptor each.
If you need more descriptors, ask the Advisory service if it is possible to have your file index increased.

(d) NO MORE SPACE FOR OUTPUT
This occurs if there is not enough room in your filespace for your output file(s).
The command `FILES ,P` will tell you your total limit and how much of it is used.
Action: DESTROY and/or ARCHIVE files to make more room. If you still have

insufficient room, contact the Advisory service to get your file limit increased.

(e) OUTPUT FILE CAPACITY EXCEEDED
This occurs if the defined maximum file size is not enough to hold your output.
Action: Either split your job up into 2 or more runs, or reDEFINE the output channel if `FILES ,P` shows your maxfilesize is greater than 1023K, which is the default.

12. Examples of running SPSSX jobs on EMAS 2900

(1) Interactively / foreground

(a) `SPSSX CONTROL=...,etc`
e.g.

```
SPSSX CONTROL=SPSSCON,LISTING=.LP15
```

(b) `OBEY filename,out`

where `filename` is a file containing the SPSSX command

```
SPSSX CONTROL=...,etc
```

and

`out` is an optional parameter giving the destination of the output.

e.g.

```
OBEY SPSSJOB1
```

(2) Batch / background

```
DETACH jobfile,cputime,paramfile
```

where
`jobfile` is a file containing the SPSSX command

```
SPSSX CONTROL=...,etc
```

`cputime` is an optional parameter giving the time limit in seconds for the whole job, default 30 seconds, and
`paramfile` is an optional parameter giving the name of a file containing document scheduling parameter assignments or .IN.

e.g. `DETACH SPSSJOB3`

e.g. `DETACH SPSSJOB4,60`

e.g. `DETACH SPSSJOB5,,.IN`
Doc param: `OUT=LP25`
Doc param: `.END`

e.g. `DETACH SPSSJOB6,90,DOCFILE`
where
`SPSSJOB6` is a file containing the SPSSX command and
`DOCFILE` contains

```
OUTNAME=JOURNAL4
.END
```

13. Examples of running SPSSX jobs on EMAS-3

(1) Interactively / foreground

(a) `SPSSX CONTROL=...,etc`

e.g.

```
SPSSX CONTROL=SPSSCON,LISTING=.LP15
```

(b) `OBEY filename,out`

where

`filename` is a file containing the SPSSX command

```
SPSSX CONTROL=...,etc
```

and

`out` is an optional parameter giving the destination of the output.

e.g.

```
OBEY SPSSJOB1
```

(2) Batch / background

```
DETACH jobfile,cputime,paramfile
```

where

`jobfile` is a file containing the SPSSX command

```
SPSSX CONTROL=...,etc
```

`cputime` is an optional parameter giving the time limit in seconds for the whole job, default 30 seconds, and

`paramfile` is an optional parameter giving the name of a file containing document scheduling parameter assignments or .IN.

e.g. `DETACH SPSSJOB3`

e.g. `DETACH SPSSJOB4,60`

e.g.
```
DETACH SPSSJOB5,,.IN
Doc param: OUT=LP25
Doc param: .END
```

e.g. `DETACH SPSSJOB6,90,DOCFILE`

where

`SPSSJOB6` is a file containing the SPSSX command and

`DOCFILE` contains

```
OUTNAME=JOURNAL4
.END
```

14. Hints on running SPSSX more efficiently

1. Use systemfiles rather than raw data.

Due both to the internal structure of SPSS-X, and to the wider range of file formats supported, raw data input is much slower than was the case with the old SPSS. Whenever a dataset is to be processed more than once, it is advisable to save it as a systemfile.

2. Minimize the number of procedure commands.

Each procedure command causes SPSS-X to read the entire dataset. For the simpler procedures (frequencies, crosstabs, etc.) reading the data is by far the most

expensive and time-consuming part of the job. Thus, for example, the commands:

```
CROSSTABS TABLES=A BY B
CROSSTABS TABLES=B BY C
```

will cost almost twice as much as the equivalent:

```
CROSSTABS TABLES=A BY B / B BY C
```

Note that, with large tables, this technique can cause SPSS-X to fail with an error message of the form 'not enough memory for this run' - in this case, it is far better to increase the available memory (using the WORKSPACE parameter of the SPSSX command) rather than splitting the command.

3. Use 'integer mode' where available.

The procedures frequencies, crosstabs and breakdown allow the user to specify the range of values which an integer-valued variable may take. This specification causes SPSS-X to use a completely different and far more efficient method of accumulating the required statistics, and should always be given when the range is known.

In this context, it is advisable to encode each variable as a small contiguous set of integers.

4. Use the DROP or KEEP subcommand with GET.

When an SPSS-X job includes more than one procedure command, the entire dataset must be written to a temporary (scratch) file by the first procedure so that it can be read again by subsequent procedures. The time required to read and write this file obviously increases with the size of the file. The use of DROP or KEEP with the GET command will result in a scratch file which includes only the variables required in the analysis. Thus, if a job uses ten variables from a file containing 100 variables, KEEPing only the required ten can make the run up to ten times faster - as well as avoiding potential problems with lack of temporary storage space.

If the same subset of variables is likely to be used in several analyses, an even greater reduction can be achieved by saving this subset as a new systemfile.

5. Organize transformation commands carefully.

As with raw data input, transformations in SPSS-X are much more flexible and powerful than with the old SPSS - and are consequently slower. It is advisable to:

a) place SELECT IF commands as early in the sequence as possible, so that other transformations are not applied to cases which will be rejected.
b) order SELECT IF commands such that those which reject the greatest proportion of the data appear first.
c) use RECODE rather than IF commands where appropriate.
d) use the FILE TYPE command rather than SELECT IF where appropriate.
e) use the functions RANGE and ANY, if appropriate, rather than complex logical expressions in IF and SELECT IF commands.

6. Avoid using permanent transformations between procedure commands

When permanent transformations occur between two procedure commands, SPSS-X must re-write the entire dataset, including the transformed variables, to a second temporary file. This is very time-consuming and can lead to difficulties in finding sufficient temporary storage space. Sequences of transformations between procedures should be preceded by the TEMPORARY command whenever possible.

7. Save transformed files for future use.

If a set of transformed variables is likely to be used in several analyses, it is far better to save a new systemfile including these variables than to perform the same transformations in each run. This also applies to the use of MATCH FILES.

8. Save correlation matrices.

Procedures regression, factor and discriminant spend much of their time calculating correlation matrices. If these procedures are used to analyse the same set of variables in different ways, it is better to save the matrix in the first run and re-read it in subsequent runs - this saves reading the entire dataset on each run.

9. Do not request unwanted statistics.

The statistics produced by procedures such as crosstabs are calculated from sums, sums of squares, cross-products, etc. which are accumulated as each case is read. This can greatly increase the time required to read the data. Most SPSS-X procedures are sufficiently 'intelligent' to avoid such calculations if the corresponding statistic was not requested. It is inadvisable, therefore, to specify STATISTICS ALL - a little extra typing can save a considerable amount of computer time.