THE SAS SYSTEM AND THE MICROCOMPUTER

P. L. Olympia, Ph.D.
Darwin Systems, Inc. and Roy F. Weston, Inc.

INTRODUCTION

At SUGI '83 I described the SAS <--> Microcomputer connection, a system that brings the power of mainframe computers available to micro users. Essentially, the system allows a user who is not necessarily computer-literate to drive SAS and SAS/GRAPH on a mainframe as a slave of the micro.

This paper discusses some of the things that we have done during the past year that may be of interest to SAS users who also are microcomputer enthusiasts. Among the topics covered are:

- Loading SAS and other mainframe data from hardcopy input
- Converting a System 2000 data base to dBASE via SAS
- Displaying graphs on a variety of devices from a single SAS/GRAPH output
- A brief survey of microcomputer statistical software
- Announcement of a remote bulletin board system to encourage the free exchange of information among SAS-Micro users.

Throughout this discussion I will be mentioning specific hardware and software that we have used. That should not be construed as endorsement of any one particular product.

LOADING SAS DATA FROM HARDCOPY

Faced with the prospect of loading mainframe data bases from scratch we have always elected to go the route of the microcomputer via formatted screens with built-in edit checks. The locally-stored data can then be uploaded to the mainframe with the aid of just about any microcomputer communications software. This process offers a number of advantages such as speed and a superior user interface.

The procedure does not work very well in those situations where the data are being collected at a large number of field offices which do not have the proper equipment. For example, we maintain a travel data base system with components in both dBASE II on a micro and SAS on the mainframe. The input data come from information that the prospective traveler types on a travel form. Because not all field offices have micros or access to the mainframe we are forced to collect the data using the world's state-of-the-art communications medium: paper. But rather than re-enter data that are already on the form we read the data directly via the DEST optical character reader hooked to the micro's communications port. The micro at this point is running a communications program, e.g., MODEM7, with its ASCII capture buffer open (via the CTRL-Y command); all data read by the DEST are simply passed to the micro and eventually stored on disk. The disk data can be edited with a text editor, e.g., WordStar, or passed to, and later edited with, dBASE II. From there, communicating the data to SAS and other mainframe software is relatively straightforward.

This approach offers many advantages:

- Paper is a universal and inexpensive medium; computer wags have always said that silicon is expensive and iron (hardware) is cheap, but nothing beats paper in terms of cost. It is also the only way to go if one has a collection of word-processors and micros with incompatible disk formats or a variety of software that store data in non-standard ways
- The DEST optical reader breaks new ground in terms of cost/benefit ratio; it is reasonably fast, supports up to 8 resident type fonts, has a low error rate, and can read, unattended, a stack of 75 documents, including xerographic copies.
- The process can be automated with the use of an operating system's batch command language and judicious programming of function keys.
- The technique is useful for electronically capturing all sorts of data that are generated at various locations, e.g., correspondence and memoranda, project abstracts and budget tables.

SYSTEM 2000 TO dBASE AND BACK VIA SAS

Many organizations have end-users running relatively small data bases on the mainframe. After discovering the versatility of micros, these same users begin to demand that the data base be off-loaded to their micros. In cases where the mainframe DBMS is heirarchical and the micro DBMS is relational, the conversion is not easily carried out without procuring specialized software.

In our case, the problem was to move an entire project tracking data base from System 2000 to dBASE II. Given the well-known constraints of dBASE II (maximum 32 fields per data base and 1000 bytes per record, no date type variables, etc.) and intrinsic differences in hierarchical and relational-type DBMS, particularly the ability of hierarchical DBMS to allow

variable-length records and string overflows, our challenge was to make the conversion using no more than two dBASE II data bases. Our solution was to use SAS to handle all interface and incompatibility problems.

Not being experts in the S2K programming language interface, we took the easy route by "unloading" the data base via a series of report writer programs driven by a TSO Clist. We then used SAS to clean the data and format it in a way that is compatible to the pre-determined structure of the dBASE II data bases. The solution around the 32 fields per data base was to pack more than one piece of information in a field, separated by a delimiter. For example, the NAMEORG field actually contained the name of the principal investigator, his phone number and organization. A dBASE command file will later take care of "unpacking" the field as needed. A compromise had to be made to handle string overflows. If a character field exceeds the predetermined length, the excess is simply discarded.

The next problem was to move the massive SAS-created data set to the micro given that the IBM mainframe is one of the worst in terms of its lack of support of even the simplest (e.g., XON/XOFF) asynchronous communications protocol. At the time, we did not have KERMIT (a versatile communications program that runs on a host of machines, e.g., TRS, APPLE, CP/M, MS-DOS, VAX, IBM CMS, etc. and whose price is right: free from Columbia University and most user groups). We conducted the transfer at 1200 bps using "blind transfer", i.e., without any error-detection protocol. Because the micro's ASCII capture buffer could hold only 42K at a time, we wrote a SAS program to segment the SAS-created data set into smaller chunks.

Once the files were on the micro's disk, loading the data base into dBASE II was a trivial task via the single command: APPE FROM filename SDF.

The entire process can of course be reversed if one is perverse enough to want to move a data base from dBASE II to S2K. The lesson to be learned from this exercise is that it is relatively easy to move data bases or files of any size or complexity from the mainframe to a micro and back if one has two things: SAS and a versatile communications program.

SAS/GRAPH PICTURES ON UNSUPPORTED DEVICES

Most people will agree that SAS/GRAPH is in a league by itself as far as graphics software go. A microcomputer user who is not fortunate enough to have any one of the many graphics devices that SAS/GRAPH supports, heretofore had only one choice: use a communications software such as TEKTERM for the Apple or PC-PLOT II for the IBM PC or XT that allows the micro console to emulate a TEKTRONIX 4010 device. Yet, when you think about it, there is a better way that takes full advantage of a micro's capability. This other poor man's alternative allows one to display the same SAS/GRAPH picture on any number of devices, both supported and unsupported, and also allows a clever user to incorporate both text and graphics in the same document.

The technique is relatively simple and, as a starting point, uses the same procedure described recently by Kelly (Kelly, 1983). Briefly, the approach consists of the following:

- Using your favorite communications program initiate a SAS/GRAPH session on the mainframe and tell SAS/GRAPH that you are using a TEK4010 device; you are lying, of course, but neither computer cares

- When SAS/GRAPH is ready to draw the picture, open your ASCII capture buffer (CA + in Crosstalk, CTRL-Y in Modem7, ESC R in BSTMS, etc) and watch the weird-looking characters show up on your screen as they are being captured by your micro, first in its RAM, and later to a disk file that you name. Figure 1 is a portion of a disk file that if "typed" to a TEK4010-emulating device will display's SAS/GRAPH's famous cowboy hat example.

- Terminate the mainframe session, and write a simple program to translate the TEK4010 commands captured on the disk file into the appropriate commands that your unsupported device(s) can interpret. More on this later.

- If your unsupported device is a letter-quality printer with graphics capability (e.g. AJ832), your conversion program should direct its output to a disk file; you can then include the graph in your document with something like WordStar's MailMerge facility.

```
VgVaY6§]¶$AF§J²]7w#]u[rYnXhXdY°]6§$AF7°JaLdO²]7s
$@c@²]7s$AeA²]7u$CgCdAc@²]7s$@uCyGwJLyN²]7w$FvHK
²]7v$DuGJvLyN²]7u$J°J²]7y%@w$]uZrYnWhWdYaZ6§]¶%@
½DG¶K§M7aPdRhSnSrRuPwMyKwIuEsA²]7s$ZoYgYcZ²]7y%@
v$]s±oZgZc±°]6¶%@²]7c%PgRoRuOvM²]6¶%K7°McOgPoPrO
vLwI²]7s%AaA²]7s%BdB²]7s%DeDcBaA²]7w%I6¶I²]7o%In
LMoP²]7g%IhLMgP²]7y%±w]u_1_±k[h[j±_°_6§±²]7v%_s&
@°@³]7k%±k&@³]6§&AD¶G²]7y%±w_v&@sA°AF6§H³]6§%±§&
@¶D½F§H7°L6§0½Q§S7°W²]7v&FyJvL°LQ6§S³]7w&JvK°K²]
7v&FHuJ°J6§H²]6§&Q¶R²]7v&QyUvW°W²]7w&UvV°V²]7v&Q
SuU°U6§S²]7v&F°F²]7v&Q°Q²]7n&FJ²]7h&FJ²]7n&QU²]7
h&QU²]7v&_y'BFwIyK²]7w'CG²]7v&_w'BvDIyK²]7r'FpDn
C1C@k&_h_j'@CaC²]7o'DdD²]7k'@D²]7r'FeFcDaC²]7u'M
sKpJdJ²]7r'KgK²]7u'MhMeKdJ²]7u'MyUwXuYrYoV1Q²]7w
'UuXrX²]7w'RvUuVpVnTkXhY°Y²]7k'VhXaX²]7n'T1UjV°V
²]6½'B§F7°JO6§T²]6¶'D§IO¶R²]6½'B¶GN½Q§T7°VY²]71'
Q°Q²]7h'QV²]7d'QV²]7y(JwGuDrCnAhAdCaD6§G¶J½NP¶U§
W7aZd[h]n]r[uZwWyUwSuOsK²]7s(DoCgCcD²]7y(JvGsEoD
```

Figure 1. A portion of TEK4010 ASCII Commands for SAS/GRAPH's cowboy hat example

To help you write your code translation program for one or more devices, you only need to understand the way a TEK4010 draws a graph. Although a detailed discussion of that is outside the scope of this paper, a few points need to be explained.

A TEK4010's way of drawing a graph is the ultimate in simplicity. All it does is move the cursor from one point to the next. If the cursor is in the "pen down" position the vector is drawn. It will draw the vector once it has received the appropriate command and it "knows" where the target coordinate is. The coordinate (X,Y) is identified by a set of four bytes; two seven-bit byte is needed to identify the X component (we call it the Hi X or HX, and the Lo X or LX bytes). Similarly, two seven-bit bytes are needed to represent the Y component (one for HY and another for LY). The vector is drawn as soon as the device receives the LX byte; if it does not receive any new HX, HY or LY byte prior to the LX, it merely uses the same bytes of the current coordinate.

How does it distinguish among the bytes or know when to move the "pen" up or down? And how does it know whether the next character is plain text or a graphics character? Simply by examining the ASCII character that it has just received. For example, the ASCII decimal 29 character is the command for "pen up". The ASCII decimal characters in the range 64-95 (that is, capital letters and a few special symbols such as @ and the curly brackets) are LX bytes. Figure 2 is a summary of the ASCII character set command grouping.

ASCII DEC	ASCII CHAR	COMMAND
26	CTRL-Z (SUB)	Clear text screen
27,12	ESC,FF	Clear text/graph screen
29	CTRL-] (GS)	Pen up
31	CTRL-_ (US)	Next char is text
32-63	numbers and special symbols	HY character
64-95	capital letters and some symbols	LX character
96-127	lowcase letters and some symbols	LY character
96-127 (LY) followed immed. by 32-63 (HY)		HX character

Figure 2. TEK4010 ASCII Commands

Clearly, for any device to interpret TEK4010 commands requires only that the device be able to respond to a pen up/down command and move the pen to a specified coordinate. This procedure allows a micro user to store one SAS/ GRAPH picture on a floppy or hard disk and replay the picture any number of times on any number of devices, both supported and unsupported, via a menu-driven code conversion program. The bonus is that it is relatively easy to integrate text and graphics in the same document.

STATISTICAL SOFTWARE FOR MICROS

While today's crop of microcomputer statistical software does not pose any serious threat to such mainframe packages as SAS and SPSS, their number and capabilities are steadily growing. The primary limitations of micro-based statistical software are the number of records and variables that they can process, the richness of the statistical procedures that they offer, and the precision of the computed values that they generate. The Platts' multiple regression benchmark data set (Figure 3) can be used to evaluate a program's computational precision. Foxwell (Foxwell, 1983) reports that with this data set SAS produces the following regression coefficients (B0 to B3): 1.53253993, 0.05525459, -2.64404468 and -0.21932197. A micro-based software should yield comparable values.

Dep. Var.	Independent Variables					
Y	X1	X2	X3	X4	X5	X6
.72	0	.44	.13	.44	.48	-8000
.75	0	.33	.12	.36	.48	0
.61	0	.49	.19	.49	.48	-45000
.14	0	.50	.13	.50	.50	-14000
.74	0	.14	.17	.25	.48	-9000
.74	0	.43	.17	.42	.48	-8000
.80	0	.37	.10	.38	.48	-25000
.16	0	.42	.08	.42	.48	-46000
.78	0	.33	.11	.34	.48	157000
.17	0	.44	.08	.44	.48	1000
.83	0	.30	.40	.44	.48	-9000
.50	0	.49	.24	.49	.49	-142000
.52	0	.47	.23	.47	.48	-36000
.71	0	.52	.14	.52	.52	-1000
.74	1	.45	.13	.45	.48	16000
.87	1	.33	.06	.33	.48	-36000
.33	1	.56	.32	.56	.56	-33000
.53	1	.52	.23	.52	.52	7000
.58	1	.34	.20	.38	.48	-3000
.57	1	.19	.21	.29	.48	6000
.00	1	.05	.00	.05	.48	-6000
.50	1	.37	.24	.37	.48	-169000
.70	1	.60	.14	.60	.60	83000
.60	1	.40	.19	.40	.48	120000
.79	1	.39	.10	.39	.48	179000
.00	1	.05	.00	.05	.48	-1000

Figure 3. Platts' Multiple Regression Data Set (as cited by Foxwell)

Figure 4 is a summary of the most popular software packages for business micros. It does not include the many programs that are written

specifically for the Apple. Most users who cannot wait for, or afford, the much-publicized IBM XT/370 will probably find either STATPAC or SYSTAT to be quite capable.

RBBS FOR SAS/MICRO USERS

To address the needs of the growing number of SAS users who are also microcomputer enthusiasts we will be initiating an experimental remote bulletin board and software exchange system around the end of April. The experiment will be evaluated at the end of four months and a decision to continue or discontinue will then be made. The service is free to anyone who has a terminal or micro and a 1200 bps modem. The system will be operated like the Remote CP/M or RBBS-PC systems around the country.

The bulletin board system will allow novice users to ask questions and advice, and expert users to provide answers. The system will have the usual public domain programs from the CP/M and MS-DOS user communities available for downloading. More importantly, we hope that SAS users who have written interesting PROCs, macros and articles on SAS will share their knowledge with others.

The system will use the XMODEM (Christiansen) protocol for transfer of binary files. Users whose communications software does not support this protocol may only transfer text files. The XMODEM protocol is supported by most communications software, particularly, MODEM7 for CP/M, CP/M-86 and MS-DOS, MITE, PC-TALK, SYSCOMM, PC-DIAL and CROSSTALK version 3.4 or later. CP/M users can download a text file which is available on the system on how to patch their PIP program to get MODEM7 for their particular machine.

The telephone number and details of operation will be made available to users (perhaps via the SUGI SIG/M Newsletter) as soon as the system is on the air.

Y'all call.

REFERENCES

1. Kelly, M., BYTE Magazine, Oct. 1983, p. 439.

2. Foxwell, H., CAPITAL PC MONITOR, Dec. 1983, p. 9

3. Direct all inquiries to: Darwin Systems, Inc. 17 Thorburn Rd., Gaithersburg, MD 20878

SOFTWARE	VENDOR	OPERATING SYSTEM	DESCRIPTION
ABSTAT	Anderson-Bell Co. P.O. Box 191 Canon City, CO 81212 (303) 275-1661	CP/M MS/PC-DOS	Fully interactive with built-in editor and good data manipulation capabilities. Allows max. of 20 variables; can share data with other software, e.g., dBASE DBF. Written in PASCAL/MT+
BMDSTAT (SIGSTAT)	Significant Statistics 3336 N. Canyon Road Provo, UT 84604 (801) 377-4860	PC-DOS (128K)	Package of 30+ pgms from BMD series, with advanced features. Max. of 80 variables. Opt. 8087 support.
MICRO DATA ANALYZER	Cambridge Information International, Inc. 238 Main St., Suite 320 Cambridge, MA 02142 (617) 354-0199	CP/M MS/PC-DOS	12-program package that handles very large data sets limited only by avail. disk space. Uses FORTRAN-like commands. Requires file sizes to be predefined. Directs output to screen or printer only.
MICROSTAT	EcoSoft 5311 N. Central Ave. Indianapolis, IN 46220 (317) 255-6476	CP/M	File-oriented package with data management subsystem for file maintenance. Written in NS BASIC, BASIC-80 and baZic
NWA STATPAK	Northwest Analytical 1532 Morrison St. Portland, OR 97205 (503) 224-7727	CP/M MS-DOS	Has good data manipulation utilities; reads ASCII data files. CP/M version written in MBASIC
STATPAC	Walonick Associates 5624 Girard Ave. Minneapolis, MN 55419 (612) 866-9022	PC-DOS MS-DOS	Emulates SPSS incl. use of control file. Handles up to 5000 cases and 255 variables. Can direct output to screen, printer or disk. Good precision. Handles both fixed and variable length file format. Written in compiled BASIC.
STATISTICIAN's MACE	Mace, Inc. 2313 Center Ave. Madison, WI 53704 (608) 241-4566	PC-DOS MS-DOS	Accepts keyboard or disk file input. Claimed to be easy to use. Evaluation copy, $30. Req. 128K
STATGRAF	Village Information Co. 55 Christopher St. New York, NY 10014 (212) 924-8314	PC-DOS MS-DOS	Interactive graphics; max 19 vars. in correlation; up to 5 levels 3-way ANOVA; can direct output to screen, printer or plotter.
SYSTAT	Systat 1127 Ashbury Ave. Evanston, IL 60202 (312) 866-5670	CP/M MS/PC-DOS	Can read triangular & hierarchical files with fixed and var. length records. Powerful file manip. ability; allows multidimensional scaling. Opt. 8087 support.
SPSS/PRO SPSS/PC	SPSS, Inc. 444 N. Michigan Ave. Chicago, IL 60611 (312) 329-2400	Dec PRO 350 PC-DOS	Has anyone seen this?

Figure 4. Brief survey of some microcomputer-based statistical software