

THE IMPLEMENTATION OF SAS VIA SUPERWYLBUR AT AIR PRODUCTS AND CHEMICALS, INC.

T. J. Cenna & T. J. Bzik

INTRODUCTION

A number of professionals at Air Products and Chemicals require the ability to conduct statistical analyses via SAS, but do not want to learn a computer language in order to do so. This paper addresses an approach to resolving that problem by making the user input as effortless as possible, while returning to him a vast amount of statistical information which he probably would not request were he coding a SAS program himself. The concept was implemented using both the text editing and remote job entry facilities available in SUPERWYLBUR, in conjunction with SAS.

WHAT SUPERWYLBUR DOES

SUPERWYLBUR is an interactive computer system marketed by Optimum Systems Incorporated (OSI). As used herein, interactive means that SUPERWYLBUR interactively builds a set of code which is then submitted as one entity to the mainframe as a batch job. While all features of SUPERWYLBUR are used at Air Products and Chemicals, only the following are of importance in the context of this presentation. SUPERWYLBUR -

1. Provides a comprehensive text editing facility at a CRT or typewriter terminal.
2. Provides the ability to create, modify, store, and retrieve text or data. This permits users to have and maintain personal databases.
3. Allows macro programming, the automatic processing of a sequence of SUPERWYLBUR commands. A SUPERWYLBUR macro, unlike a SAS macro, is an interactive program.
4. Allows remote job entry and retrieval from the computer's normal batch stream.

These features can be integrated with SAS as described in Figure 1.

The user, while logged onto the SUPERWYLBUR interactive computer system, calls up a SUPERWYLBUR macro program by the use of a keyword. This interactive program then queries the user in English text. Based on responses to these queries, the macro copies portions of a candidate set of SAS code and merges this code with a data file to form a temporary 'SAS' file. This temporary 'SAS' file contains an executable SAS program which is automatically submitted to the mainframe computer. When the job has been executed, it can be previewed at the CRT via SUPERWYLBUR.

A GENERAL PURPOSE PROGRAM

In our experience we have found that a general purpose SUPERWYLBUR macro program named QWIKSTAT can handle about 95% of the needs of our user community. By the use of a simple keyword structure, QWIKSTAT allows the user to directly or indirectly access numerous SAS procedures (ANOVA, CHART, CORR, GLM, KSLTEST, MATRIX, MEANS, PLOT, PRINT, RANK, RSQUARE, SORT, STANDARD, and STEPWISE). In a given run, any number of keyword requests can be made, including repeat requests of the same keyword. Prior to reaching the keyword stage in QWIKSTAT the user can, at his option, restructure his data base via any desired transformations. QWIKSTAT was designed with the specific purpose of giving the professionals at Air Products and Chemicals the ability to conduct general purpose statistical analyses. For users who require other SAS procedures, there is an option which allows a preconstructed file of SAS source code to be appended to the job. Figure 2 shows the prompts and responses involved in a limited session of QWIKSTAT.

CUSTOMIZED PROGRAMS

Several customized programs have been created to support particular needs of the user community in Air Products and Chemicals. Many of these programs generate various Quality Analysis Reports. These programs create special analyses and report formats, but at the price of being much less general in their capabilities. A complete listing of the SUPERWYLBUR source code for a special purpose macro program is given in Figure 3. This program allows the user to plot his data. This simple example was constructed since an actual quality analysis report could contain proprietary information.

ADVANTAGES

Several advantages are realized or can be realized by the use of the previously described system:

Ease of Use

Use of these programs requires minimal user effort and virtually no user training while providing useful analysis capabilities. The fact that the queries and responses are in English help to create a "friendly" environment for the user. Each program can be self-documenting in that it can contain all necessary user instructions.

Statistical Education/Guidance

One of the more important purposes of devising this system is the opportunity it presents to further the statistical education of the user. Our user community has many members who have had minimal training in statistics. As statisticians we feel responsible to make sure the user has a complete analysis tool while having any major limitations of that particular tool pointed out to him. The philosophy of QWIKSTAT is to automatically provide diagnostics and interpretive comments to aid the user in deciding whether or not a particular statistical technique is appropriate.

For example, it is well known that regression analysis is one of the more misunderstood and abused techniques of statistical analysis. To help address this, QWIKSTAT, which allows multiple regression analyses, automatically includes seven pages of diagnostics associated with the regression. Included with the diagnostics is a complete analysis of the residuals.

When a stepwise regression is requested, the user is automatically provided with a forward stepwise regression, MAXR, and RSQUARE. By providing all this information we hope to dispell the notion that stepwise regression provides a unique model and to stimulate questions on the topic from our users. When using the MAXR option, the user is warned that the "best" n-term model in MAXR is only the best model found according to the search algorithm. The algorithm could fail to locate the true "best" n-term model for $n > 2$; RSQUARE is a valuable aid in this respect. In fact, we have demonstrated at least one case where MAXR found the "best" two variable model to have an r-square=.23 while a two-term model with r-square=.98 was not located.

To a limited extent, QWIKSTAT has the ability to provide direction for the analysis through the types of queries given. For example, when performing a multiple regression analysis, the user will receive the options to plot the residuals against any independent variables or a time variable.

Error Protection

Good macro program logic can eliminate many sources of error. Errors in SAS logic and punctuation are practically nonexistent due to the structured way in which the SUPERWYLBUR macro generates the source code. Simple typographical errors to queries (invalid responses or addressing of nonexistent data bases) are often detectable by the SUPERWYLBUR macro and the query is then repeated. In general a SUPERWYLBUR macro can be made relatively foolproof without excessive programming effort.

Modular Construction

Each of the SUPERWYLBUR macros has been constructed in a modular fashion. An excellent example of this is found in QWIKSTAT where the keyword structure defines the major program modules. The flow chart of QWIKSTAT's modular construction is given in Figure 5. This permits simpler logic, quicker debugging, and easier installation of new program features if any are desired.

Software Requirements

This interactive system as described herein, will offer some software advantages when contrasted to interactive SAS. If a current computer systems hardware-software configuration doesn't have either TSO (time sharing option) or CMS available, this interactive system is still feasible, whereas interactive SAS isn't. For our computer configuration, the interactive system described here also places a smaller demand on computer system resources.

Productivity Aid

These macros can be viewed as a productivity aid, even for the experienced SAS programmer, since a large set of SAS source code can be generated via a minimal number of prompts and responses. The system described herein can yield significant time savings in terms of "people costs" when contrasted to an entirely interactive system. This is because the user can leave the CRT while the SAS job is being executed, return later, and not suffer many of the consequences of slow computer time. Also, if a large analysis is being performed, the interactive system described herein does not require the user to watch the job execute while at the CRT. Note that the previous two advantages mentioned above are simply potential batch mode advantages.

SUMMARY

The system is easy to program and maintain while giving users useful analysis capabilities on personal databases. This system has resulted in a satisfied user group. Often, members of the user community will request revisions which they consider useful. The statistical education and guidance we offer through and because of this system often spins off into consulting work.

Figure 1 - The SUPERWYLBUR/SAS Connection

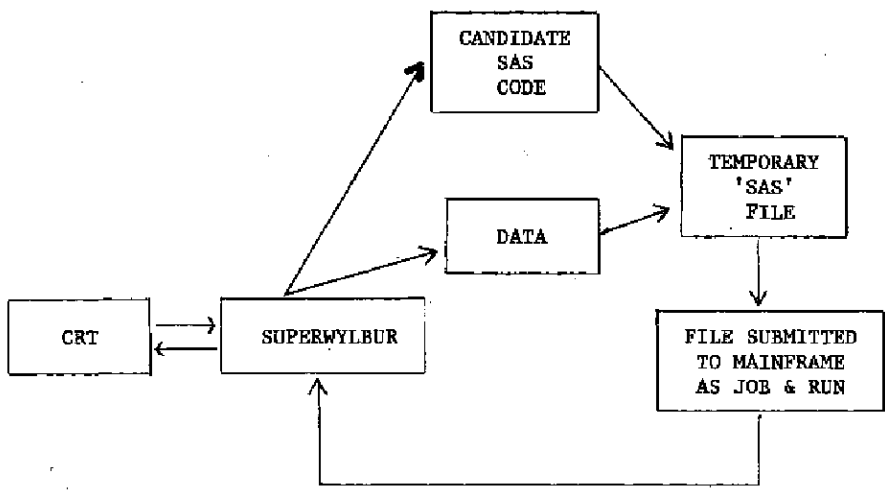


Figure 2 - Sample Session with QWIKSTAT

In this sample session the user is already logged onto SUPERWYLBUR and has already begun the session with QWIKSTAT. QWIKSTAT was accessed by simply typing 'qwikstat'. User entries are in small letters and computer responses in capital letters.

NAME OF DATA FILE:(EX. DATASETS:DATA)? datasets:data

DO YOU WANT TO TRANSFORM ANY OF YOUR VARIABLES OR CREATE NEW VARIABLES? no

AT THIS POINT, YOU CAN DO A VARIETY OF THINGS WITH YOUR DATA BY TYPING ONE OF THESE KEYWORDS

KEYWORD	OPERATION
REG	MULTIPLE AND STEPWISE REGRESSION
PLOT	GRAPH VARIABLE IN TWO DIMENSIONS
DESCRIBE	BAR CHARTS, MEANS, SUMMARY STATISTICS
ANOVA	ANALYSIS OF VARIANCE FOR RANDOMIZED COMPLETE BLOCK DESIGN (MAIN AND CROSSED EFFECTS).

KEYWORD= ? plot

VARIABLE TO BE PLOTTED ON X AXIS IS? a

VARIABLE TO BE PLOTTED ON Y AXIS IS? b

ANOTHER PLOT? no

KEYWORD OPERATION COMPLETE. DO YOU WANT ANOTHER KEYWORD? no

1596 IS YOUR JOB NUMBER

At this point the job has been submitted to the mainframe computer and the session with QWIKSTAT is complete.

Figure 3 - Source Code from a Simple SUPERWYLBUR Macro

```

*** THE SIMPLE SUPERWYLBUR MACRO ILLUSTRATED HERE HAS HAD ALL ***
*** REFERENCES TO LINE NUMBERS REPLACED BY ILLUSTRATIVE STATEMENT ***
*** LABELS. THESE STATEMENT LABELS WOULD HAVE TO BE REPLACED BY ***
*** LINE NUMBERS BEFORE EXECUTING THIS MACRO. ***
*** T SAMP IS THE TEMPORARY FILE IN WHICH THE SAS SOURCE CODE ***
*** IS BEING COLLECTED. ***
SET EXIT ATTENTION @EXIT
SET EXIT ERROR @EXIT
TYPE ' SAMPLE MACRO FOR PLOTS '
TYPE ' HITTING THE BREAK OR ATTENTION KEY WILL ALLOW YOU TO EXIT FROM '
TYPE ' THIS PROCEDURE. '
TYPE
TYPE ' YOU WILL NEED TO INPUT SEVERAL PIECES OF INFORMATION '
TYPE ' (1) A JOBCARD FILE '
TYPE ' (2) THE APPROPRIATE DATA FILE '
TYPE ' (3) AN OPTIONAL TITLE OF YOUR CHOICE '
TYPE
TYPE ' A FILE CONTAINING A CLASS H JOBCARD IS REQUIRED. HIT THE '
TYPE ' ATTENTION OR BREAK KEY AND SET ONE UP IF YOU DO NOT HAVE ONE. '
*** JOBCARD FILE ENTRY SEGMENT ***
SET EXIT ERROR @JOBCARDFILE
@JOBCARDFILE DEMAND STR JC 'NAME OF YOUR JOBCARD FILE = ?'
QT COPY ALL FR %JC TO 1 T SAMP NEW
SET EXIT ERROR @EXIT
*** COPY SAS SOURCE CODE AT END OF MACRO TO TEMPORARY FILE SAMP ***
QT COPY @SASCODR1B/@SASCODE1E T %.MACRO TO END T SAMP
*** DATA FILE ENTRY SEGMENT ***
SET EXIT ERROR @DATAFILE
@DATAFILE DEMAND STR DC 'NAME OF THE APPROPRIATE DATA FILE = ?'
QT COPY ALL FR %DC TO END T SAMP
SET EXIT ERROR @EXIT
*** OPTIONAL TITLE ENTRY SEGMENT ***
TYPE ' YOU CAN CHOOSE A TITLE TO BE PRINTED WITH ALL THE PLOTS. '
TYPE ' TYPE THE TITLE OF YOUR CHOICE AFTER THE PROMPT "TITLE = ?". '
TYPE ' IF NO OPTIONAL TITLE IS DESIRED SIMPLY PRESS THE ENTER BUTTON '
*** SAS MACRO FOR OPTIONAL TITLE IS BUILT HERE. ***
SET EXIT ERROR @TITLEIN
@TITLEIN DEMAND STR TITLEINP 'TITLE = ? '
IF (TITLEINP = '') LET TITLEINP = ' '
LET BLD = 'MACRO T11 TITLE1 '
LET LAST = '; %'
LET BLD = %%BLD %%TITLEINP %%LAST
LET .LINE: (.END: ('SAMP'); 'SAMP') = %%BLD
*** COPY SAS SOURCE CODE AT END OF MACRO TO TEMPORARY FILE SAMP ***
QT COPY @SASCODE2B/@SASCODE2E T %.MACRO TO END T SAMP
*** PLOT SPECIFICATION ENTRY SEGMENT ***
TYPE ' ENTER THE PLOTS YOU WANT, ONE LINE AT A TIME, IN THE FORM A*B '
TYPE ' WHERE A IS TO BE PLOTTED ON THE VERTICAL AXIS. IF THERE ARE NO '
TYPE ' MORE PLOTS, ENTER THE WORD "DONE". '
@PLOT PROMPT DEM STR GPH '?'
IF (GPH='DONE') GO @NOMOREPLOTS
QT COP @SASCODE3 T %.MACRO TO END T SAMP
QT CH '$GPH' TO %%GPH IN T SAMP
GO @PLOT PROMPT
*** JOB SUBMITTED HERE. ***
@NOMOREPLOTS RUN UNN FETCH T SAMP
*** DELETE TEMPORARY FILE SAMP HERE. ***
@EXIT CLEAR T SAMP
*** SUPERWYLBUR MACRO EXITED HERE. ***
RETURN
@SASCODE1B //SAS EXEC SAS
DATA PLOTS;
INPUT A B C D E F;
@SASCODE1F CARDS;
@SASCODE2B PROC PLOT;
@SASCODE2E T11;
@SASCODE3 PLOT $GPH;

```

Figure 4 - QWIKSTAT's Modular Construction

