USING CMS SAS IN CONJUNCTION WITH IBM'S SYSTEM R RELATIONAL DATA BASE MANAGEMENT SYSTEM

T. J. Vidmar and J. P. Maile, The Upjohn Company, Kalamazoo, MI USA

One of the research units at The Upjohn Company experienced a problem with the large amount of time it took to analyze data generated from their research animals. In order to reduce their dependence on the statisticians and programmers a CMS EXEC was written to allow them the ability to process and analyze the data in a matter of minutes.

The user need not have a knowledge of the workings of IBM's System R Relational Data Base Management System or knowledge of writing SAS programs. The user must, however, be familiar with some basic CMS commands, know how to use the Upjohn full screen editor which is available for editing CMS files and also the ability to interpret the statistical output.

The "real" system begins with the drawing of a blood sample from a research animal. This sample is labeled with the animal identifier number, the experiment number, and other information relevant to the study. The sample is then sent to the Upjohn Clinical Research Laboratory for analysis. Upon completion of the analysis, laboratory personnel enter the results into a System R Data Base which resides on an IBM 370/148 computer. A hard copy report listing the results for each research animal is generated and sent to the researcher.

An abbreviated example of the animal data in the laboratory data base can be seen in Figure 1. This listing includes only those values which are of interest to the researcher; all other values are omitted.

To use the EXEC, the user must first log to his/her VM computer. After the logging procedure, the user enters "RATS", the name of the CMS EXEC which was created for the user. This EXEC uses terminal prompting to guide the user through the various routines. A few system messages will be displayed on the terminal, followed by the question, "DO YOU WANT TO RETRIEVE A NEW BATCH OF DATA? (Y/N):". If the user enters "Y", the old CMS data file will be erased and the terminal will prompt with "ENTER EXPERIMENT NUMBER". After the user enters the experiment number, a Systeam R SQL Language Cursor referencing the experiment number is opened and the data is retrieved and written to a CMS file with a file name of RATDATA and a file type of DATA on the user's A disk. A message is displayed on the terminal specifying the number of records retrieved. If the record count is equal to 0, a return code of 64 is returned to the EXEC. It should be noted that each time a new data set is retrieved, the old CMS file is erased and a new file created. This is to prevent the user from accessing data which belongs to a different experiment.

If the EXEC receives a return code of 64, the following message is displayed: "IN TRYING TO RETRIEVE DATA FROM THE CRL DATA BASE, NO RECORDS WERE FOUND. THE RUN HAS BEEN ABORTED". The EXEC will then terminate. If the EXEC receives a return code of 0 or if "N" was entered when the terminal asked whether or not a new batch of data was to be retrieved, the terminal will display the following message: "DO YOU WISH TO EDIT THE RETRIEVED DATA? (Y/N):". A response of "Y" lets the user examine and possibly correct the data by way of the editor function.

Assuming the user has examined and corrected the data, the EXEC invokes a PL1 program. This program queries the user as to which treatment groups should be eliminated (if any) from the statistical analysis. The program next reads in a SAS program and searches for an insertion marker. If the user specified that no groups should be eliminated, a blank line is inserted. Otherwise, the program inserts a SAS statement like the following: "If GP=1|GP=2|...|GP=N then delete;", corresponding to the N groups the user wishes to eliminate.

The modified SAS program is then executed. The CMS file containing the retrieved and user verified data is read as input. A summary table like the one in Figure 2 is produced. The data for each assay is then ranked and a table like the one in Figure 3 is produced. Finally, an Upjohn written procedure using the ranked data is used to perform a one-way analysis of variance. A sample of this output may be found in Figure 4.

After the SAS program runs the statistical analysis, the user is given the option to view the output of that program using the editor function. If hard copy is desired, that option is also given to the user. He or she simply responds "Y" to the question: "WOULD YOU LIKE A PRINTOUT FOR THIS RUN? (Y/N)".

During execution of the EXEC, the console is spooled to the user's reader. This causes all terminal input, messages and diagnostics, to be written to the user's reader. If the run aborts for any reason which the user cannot understand, the programmer need only access this reader file to determine exactly the reason for termination.

Acknowledgements

The authors wish to thank Patty Kemp, Bob Rush, Mary Ruwart, Bob Simmons, and Tim Tesar of The Upjohn Company for their efforts.

FIGURE 1
SEGMENT OF LABORATORY DATABASE

UNO	DRND	INV	SBJ	MIS	TST	QNTVAL
RATS	2444563	9009	101	15	1098	0.6000
	2444563	9009	101	15	166	6620.0000
	2444563	9009	101	15	167	1155.0000
	2444563	9009	101	15	189	?
	2444563	9009	101	15	416	500.0000
	2444563	9009	101	15	728	1.3000
	2444563	9009	101	15	729	2.4000
	2444563	9009	101	15	1310	153.0000

Figure 2
SUMMARY STATISTICS BASED ON THE RAW DATA
EXPERIMENT #15

	CREAT	SGOT	SGPT	ALK P.	BILI	URIC	GLU	BUN
GROUP=1								
MEAN	0.617	4911.250	983.13	327.167	0.933	2.250	143	24.750
S.E.	0.017	1438.7	242	41.935	0.214	0.056	5.145	1.877
N	6	8	8	6	6	6	6	6
GROUP=2								
MEAN	0.583	4034.375	763.13	463.000	0.650	1.750	130.67	24.929
S.E.	0.017	797.11	145.27	85.513	0.096	0.112	2.951	1.787
N	6	8	8	6	6	6	6	7
GROUP=3								
MEAN	0.543	2870.000	666.88	304.857	0.486	1.771	134.14	27.757
S.E.	0.030	548.37	97.110	26.182	0.040	0.108	3.232	1.450
N	7	8	8	7	7	7	7	7
GROUP=4								
MEAN	0.700	7575.714	1622.1	375.250	1.225	2.125	130	30.057
S.E.	0.041	1966.1	375.24	41.536	0.293	0.202	5.986	2.216
N	4	7	7	4	4	4	4	7
GROUP=5								
MEAN	0.540	4323.333	1002.5	348.000	0.680	1.680	133.8	24.680
S.E.	0.024	601.18	197.76	11.777	0.049	0.198	2.417	1.235
N	5	6	6	5	5	5	5	5

Figure 3
SUMMARY STATISTICS FOR EACH GROUP USED IN THE ANALYSIS BASED ON THE RANK TRANSFORMED DATA EXPERIMENT #15

	CREAT	SGOT	SGPT	ALK P.	BILI	URIC	GLU	BUN
GROUP=1								
MEAN	17.667	18.625	19.938	12.500	17.333	23.083	20.000	13.333
S.E.	1.667	4.721	4.272	4.054	3.561	1.114	3.751	4.139
N	6	8	8	6	6	6	6	6
GROUP=2								
MEAN	14.083	18.250	15.563	18.917	13.250	10.583	10.917	13.714
S.E.	1.917	3.437	3.741	3.597	2.851	2.959	2.856	4.065
N	6	8	8	6	6	6	6	7
GROUP=3								
MEAN	10.571	13.500	15.000	9.714	6.714	11.786	14.571	20.643
S.E.	2.597	3.343	2.671	2.818	2.222	2.198	3.328	3.474
N	7	8	8	7	7	7	7	7
GROUP=4								
MEAN	24.000	24.786	25.000	17.125	22.750	18.625	12.125	21.786
S.E.	2.708	4.882	5.076	3.665	2.976	3.544	4.520	2.777
N	4	7	7	4	4	4	4	7
GROUP=5								
MEAN	9.100	21.083	20.667	16.200	16.900	9.400	14.000	11.000
S.E.	2.817	2.403	3.790	2.311	1.470	4.334	2.828	2.297
N	5	6	6	5	5	5	5	5

Figure 4
AOV — RANK TRANSFORM

CREATININE
EXPERIMENT #15

PROCEDURE AOVMEAN (VERSION OF 03/14/79)

DATA: GROUP	N	MEAN	S.D.
1	6	17.667	4.082
2	6	14.083	4.695
3	7	10.571	6.870
4	4	24.000	5.416
5	5	9.100	6.299
TOTAL	28	14.500	

ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F	PROB>F	VARIANCE
BETWEEN GROUPS	4	676.04	169.01	5.373	0.0033	24.769
WITHIN GROUPS	23	723.46	31.45			
TOTAL	27	1399.50				

PAIRWISE COMPARISONS

GROUPS	N1	N2	DIFFERENCE	S.D.	T-VALUE	PROB>ABS(T)
1, 2	6	6	3.5833	3.2380	1.1066	0.27989
1, 3	6	7	7.0952	3.1202	2.2739	0.03262
1, 4	6	4	-6.3333	3.6202	-1.7494	0.09355
1, 5	6	5	8.5667	3.3961	2.5225	0.01902
2, 3	6	7	3.5119	3.1202	1.1255	0.27198
2, 4	6	4	-9.9167	3.6202	-2.7392	0.01169
2, 5	6	5	4.9833	3.3961	1.4674	0.15582
3, 4	7	4	-13.4286	3.5153	-3.8201	0.00088
3, 5	7	5	1.4714	3.2840	0.4481	0.65830
4, 5	4	5	14.9000	3.7623	3.9604	0.00062

NOTE: THE PROBABILITIES PRINTED ARE FOR THE PAIRWISE COMPARISONS AND DO NOT REPRESENT EXPERIMENTWISE ERROR RATES.

BARTLETTS TEST FOR HOMOGENEITY OF VARIANCES:
CHI-SQUARE = 1.647 D.F. = 4
PROB. > CHI-SQUARE = 0.8003