THE MACROPEDIA: MACROS FOR LISTING AND SUMMARIZING DATA FROM CLINICAL TRIALS

Diane Wold, Burroughs Wellcome Company
Paul A. McSorley, Burroughs Wellcome Company
Sue Allen, Burroughs Wellcome Company
Janna D. Young, Burroughs Wellcome Company

Abstract

The Macropedia, written by members of the Burroughs Wellcome Company Clinical Statistics Department, is a group of macros which handle routine programming tasks. The first of three finished macros prints summary statistics, by treatment group, for any number of numeric or categorical variables. The others handle clinical laboratory data. One prints results and attaches flags to values which are above or below normal ranges. The second prints and plots summary statistics by time. The programs allow the user to specify labels and formats as well as variables for analysis, and produce output suitable for inclusion in final reports. This flexibility is achieved by having the user construct a data set which includes formats, labels, and other information about the variables involved in the analysis. The name of this special data set is a parameter of the macro, which transfers the information to macro variables. To aid the user in constructing these special data sets, interactive entry screens were created which are accessed using CMS EXEC's. The user inputs and manipulates the data to be summarized or listed as usual, then includes a line of code which imports the macro, another which imports the special data set, and, finally, the macro call.

Introduction

The Macropedia was developed by members of a Quality Circle in the Clinical Statistics Department of Burroughs Wellcome Co. We undertook this project to teach ourselves the SAS® macro language and to develop macros which could be used to perform some of the standard, repetitive, programming tasks facing members of our department.

Three macros have been completed. One, called SUMSTAT, produces summary statistics, by treatment, of any number of numeric or categorical variables. LABLIST produces listings of clinical laboratory results which include flags for values which are above or below normal ranges. SUMPLOT calculates summary statistics for a number of time points, prints them, and plots them.

Summary Statistics Macro

Figure 1 is a flowchart of the program and Figure 2 shows a page of sample output. The complete output consists of a page for each center. Some variables are summarized “numerically,” with mean, standard deviation, etc., and some are summarized “categorically,” with frequencies and percents.

SUMSTAT requires a good deal of information from the user. The name of the data set to be analyzed, a list of variables to be summarized, and an option for including a “total” column in the output are parameters for the macro. The remainder of the information is put into a special SAS data set whose name is a parameter for the macro. The default value for the name is VBLES, so we generally refer to this data set as the VBLES data set. Figure 3 shows the VBLES data set used to create the sample output. For each variable, the type, variable name, format, and label are given. “Type” specifies whether a variable is a subsetting variable, like center in the example, a treatment-type variable, or a variable to be summarized.

Laboratory Listing Macro

Figure 4 is a flowchart for LABLIST while Figure 5 shows a page of sample output. This page is part of a listing for one quarter. The data are sorted by patient, treatment, and date. Patient and treatment are used with both BY and ID statements in PROC PRINT. High and low flags (H and L) are attached to values above and below normal ranges. The name of the input data set, a list of lab variables to be printed, and the name of the VBLES data set are the parameters for LABLIST. The VBLES data set for LABLIST includes the same kinds of information as the VBLES data set for SUMSTAT except that there are four “types.” A variable can be a subsetting variable, like quarter in the example; a by-id variable like patient and treatment in the example; a by variable, like date in the example; or a “lab” variable, like study day and all the labs in the example.

The LABLIST VBLES data set also contains information on normal ranges. If only one normal range is to be used, then a high and low are specified. If more than one normal range is to be used for a particular lab, the user specifies a “split variable,” a variable whose value is to be used to choose the normal range.

Laboratory Summary and Plotting Macros

SUMPLOT is a group of macros built around a core macro, LABCOM, which manipulates the data, fits it into time windows, and calculates the summaries. The flowchart in Figure 6 shows how the printing and plotting macros relate to LABCOM. The plotting macros, SEPLOT and QPLOT, have the same structure, so only SEPLOT is shown. The printing macro, RAWSUM, sorts the summaries.
in two different orders and produces listings for each order. Samples of these listings are shown in Figures 7 and 8. Figures 9 and 10 show outputs from SEPLOT, a graph of mean plus and minus standard error, and QPLOT, a quartile plot. In both types of plots, summaries for different treatments at one time are "offset," or moved slightly to the side, so that they are not plotted on top of each other.

The parameters of RAWSUM, SEPLCT, and QPLOT include the name of the input data set, a list of labs to be summarized, and the name of the VBLES data set, which contains information on patient, time, treatment, and laboratory variables. Information on time includes ranges and target times for the time windows. For SEPLOT or QPLOT, the VBLES data set also includes information needed for the axes of the plots.

Entry Screens

Data entry screens were developed to aid in the creation of the VBLES data sets required by the macros. The screens were developed in two parts. First, the screens or panels were developed using the IBM Virtual Machine/System Product (VM/SP) Display Management System for CMS. During this phase, size, color, and input field widths were set. Next, CMS EXEC's were written to call and drive the panels. These EXEC's were written in the IBM VM/SP Restructured Extended Executor (REXX) language.

Each EXEC builds a CMS file which contains code for a DATA step and "cards" containing data. The cards contain the data from the entry screen and the DATA step inputs the cards and creates the VBLES data set.

There is a main menu, MACROPED, from which one can move to panels for SUMSTAT, LABLIST, TIME, RAWSUM, or PLOT. The LABLIST and TIME panels call secondary panels which provide for entry of information on lab normals and on time windows. A feature of the RAWSUM and PLOT EXEC's is that they can recall information originally entered for LABLIST so that information which is similar for the two macros (variable names, formats, labels) need not be reentered.

When the user first enters the MACROPED main menu, he or she must specify the filename and may specify the filetype of a CMS file (the default filetype is SAS). The EXEC opens the specified CMS file or, if it does not exist, creates it. If a new file is created, appropriate SAS input code is read into it. Each time the user completes an entry screen, the data are transferred to the CMS file, ready to be read as cards by the input code. The EXEC's include some simple error checks and messages to ensure that the input data is in the format expected by the macros.

LABLIST and the various parts of SUMPLCT share one CMS file with a DATA step which creates data sets called VBLES, TVBLES, RVBLES, and PVBLES. All of these are referred to as "VBLES" data sets. Typically, cards for the LABLIST VBLES data set (VBLES) are created first. Later, when the user needs RAWSUM or one of the plot macros, cards for the appropriate VBLES data sets are added.

The User's Perspective

The user must create a VBLES data set. This step replaces the specification of variables, labels, formats, time windows, and axis information which would be needed if the data were processed without the macros.

The user must create a data set containing the data to be analyzed. This step is the same as it would be if the data were processed without the macros.

The user must include the macro code and the VBLES file and call the macro. We operate under two different systems, MVS and CMS. For MVS programs, $GET statements are used to include the macro and the VBLES file, while %INCLUDE statements are used in CMS programs. Below are typical statements which prepare for and call LABLIST. They assume that the lab data are in a data set called LABDATA and that the VBLES data set is named LABVARS SAS. HGB, HCT, etc. are the names of the lab variables to be printed. The name of the VBLES data set does not appear since this is a keyword parameter with a default value.

%INCLUDE LABLIST;
%INCLUDE LABVARS;
%LABLIST (LABDATA,HGB HCT WBC LYMPH)
GRAN MONO EOS BASO PLATE)

Conclusion

The three completed Macropedia macros provide standard tables, listings, and plots easily. Through the use of a special SAS data set, the "VBLES" data set, the user specifies labels, formats, and other information so that the output from the macros is "customized" and ready for inclusion in final reports.

The data we work with varies from project to project, and we have, in the past, written programs specific to particular projects. The SAS macro language has allowed us to develop "production" software which retains the flexibility we need.

Acknowledgement

This project was an undertaking of the entire Mean Squares Quality Circle. It could not have been accomplished by the authors alone, and they wish to acknowledge the efforts of the other members of the circle, John Horton, Lena Hollman, Anne Clemmer, George DeMuth, Sandy Brendell, Lynn Dix, and Gail Rogers.

Notes:

1. SAS is a registered trademark of SAS Institute Inc., Cary NC, USA.
2. This paper describes a collection of macros that members of the Burroughs Wellcome Department of Clinical Statistics have used to summarize and list data from clinical trials. There is no commitment on the part of Burroughs Wellcome or the Institute to support or distribute this software.

3. Copies of the code for the programs described may be requested by writing to

Diane Wold
Clinical Statistics Department
Burroughs Wellcome Company
3030 Cornwallis Road
Research Triangle Park, NC 27709

FIGURE 4

Transfer information from VIBLES to macro variables. Count lab variables.
Create blanks. Count subs. Set k=1
Call FLPRT for kth subset
Start FLPRT
Call FLPRT for entire set
Flag all lab variables:
[i=1]
[i=i+1]
Last lab variable?
N Y
Sort, using two groups of sort variables. Print using first group of sort variables in BY and ID statements.
End FLPRT
Last subset?
N Y
End LABLIST

FIGURE 1

Start SUMSTAT
Count subsets. Make subsets. Set k=1
Y k=k+1 N
Call SUMMARY for kth subset
Call SUMMARY for entire set
Y Create _DATA1, data set with recorded treatment variable.
Y TOT=Y? N
Create _DATA2, data set with dummy treatment variable identically equal to 1.
N th variable categorical?
N Y
Y Treatment Variable?
N Y
Y Treatment Variable?
N Y
Call NUM for ith variable in _DATA1
Call NUM for ith variable in _DATA2
Call CAT for ith variable in _DATA1
Call CAT for ith variable in _DATA2
Y TOT=Y? N
Y Combine output for _DATA2 with output for _DATA1
Y Combine output for _DATA2 with output for _DATA1
Y First variable?
N Y
Y Last variable?
N Y
End SUMMARY
Print summary table
N Last subset?
N Y
End SUMSTAT
FIGURE 6

Call LABCOM

Use PDF function to create character variable which contains values of summary statistics in format specified for lab.

Set by treatment, time, window, lab.

Print using treatment and time window in BY statement in ID statement.

Set by lab, treatment, time window

Print using treatment and time window in BY and ID statements.

Call LABCOM

FIGURE 5

SAMPLE OUTPUT FROM THE LABLIST MACRO

CLINICAL CHEMISTRY

Quarter = 1st

<table>
<thead>
<tr>
<th>Patient</th>
<th>Treatment Group</th>
<th>Date</th>
<th>Study day</th>
<th>Creatinine (mg/dl)</th>
<th>BUN (mg/dl)</th>
<th>Sodium (mEq/L)</th>
<th>Potassium (mEq/L)</th>
<th>Chloride (mEq/L)</th>
<th>Bilirubin (mg/dl)</th>
<th>SGOT (U/L)</th>
<th>Alk. Phos (U/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>101</td>
<td>A</td>
<td>11/05/86</td>
<td>8</td>
<td>0.5</td>
<td>11</td>
<td>136</td>
<td>4.1</td>
<td>107</td>
<td>24</td>
<td>0.4</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/20/86</td>
<td>23</td>
<td>0.8</td>
<td>13</td>
<td>134 L</td>
<td>3.8</td>
<td>105</td>
<td>23</td>
<td>0.3</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12/05/86</td>
<td>38</td>
<td>0.7</td>
<td>8</td>
<td>137</td>
<td>3.9</td>
<td>106</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>12/22/86</td>
<td>55</td>
<td>0.7</td>
<td>5 L</td>
<td>137</td>
<td>3.8</td>
<td>103</td>
<td>27</td>
<td>0.2</td>
<td>22</td>
</tr>
<tr>
<td>101</td>
<td>B</td>
<td>10/28/86</td>
<td>0</td>
<td>0.6</td>
<td>14</td>
<td>134 L</td>
<td>3.9</td>
<td>107</td>
<td>24</td>
<td>0.5</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/13/86</td>
<td>16</td>
<td>0.4</td>
<td>10</td>
<td>135</td>
<td>4.0</td>
<td>101</td>
<td>26</td>
<td>0.3</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/28/86</td>
<td>31</td>
<td>0.6</td>
<td>5 L</td>
<td>135</td>
<td>4.0</td>
<td>106</td>
<td>25</td>
<td>0.4</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12/11/86</td>
<td>44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>01/05/87</td>
<td>69</td>
<td>0.6</td>
<td>11</td>
<td>135</td>
<td>4.0</td>
<td>104</td>
<td>25</td>
<td>0.7</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01/20/87</td>
<td>84</td>
<td>0.5</td>
<td>8</td>
<td>132 L</td>
<td>4.1</td>
<td>103</td>
<td>25</td>
<td>0.7</td>
<td>27</td>
</tr>
<tr>
<td>102</td>
<td>A</td>
<td>11/05/86</td>
<td>0</td>
<td>0.5</td>
<td>11</td>
<td>143</td>
<td>4.1</td>
<td>110 H</td>
<td>24</td>
<td>0.5</td>
<td>85 H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/20/86</td>
<td>15</td>
<td>0.5</td>
<td>8</td>
<td>138 L</td>
<td>2.6</td>
<td>106</td>
<td>22</td>
<td>1.0</td>
<td>58 H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12/04/86</td>
<td>29</td>
<td>0.6</td>
<td>7</td>
<td>135</td>
<td>3.9</td>
<td>102</td>
<td>25</td>
<td>0.3</td>
<td>48 H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12/30/86</td>
<td>55</td>
<td>0.5</td>
<td>11</td>
<td>136</td>
<td>4.4</td>
<td>105</td>
<td>24</td>
<td>0.3</td>
<td>60 H</td>
</tr>
<tr>
<td>102</td>
<td>B</td>
<td>11/13/86</td>
<td>8</td>
<td>0.6</td>
<td>8</td>
<td>137</td>
<td>3.4</td>
<td>104</td>
<td>25</td>
<td>0.6</td>
<td>85 H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11/29/86</td>
<td>24</td>
<td>0.4</td>
<td>12</td>
<td>135</td>
<td>3.6</td>
<td>104</td>
<td>26</td>
<td>0.6</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td></td>
<td>12/18/86</td>
<td>43</td>
<td>0.5</td>
<td>10</td>
<td>138</td>
<td>4.0</td>
<td>106</td>
<td>24</td>
<td>0.4</td>
<td>45 H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01/13/87</td>
<td>69</td>
<td>0.4</td>
<td>14</td>
<td>139</td>
<td>3.8</td>
<td>105</td>
<td>25</td>
<td>0.4</td>
<td>80 H</td>
</tr>
<tr>
<td></td>
<td></td>
<td>01/28/87</td>
<td>84</td>
<td>0.4</td>
<td>8</td>
<td>141</td>
<td>3.7</td>
<td>107</td>
<td>25</td>
<td>0.3</td>
<td>53 H</td>
</tr>
<tr>
<td>103</td>
<td>A</td>
<td>01/19/87</td>
<td>0</td>
<td>0.5</td>
<td>12</td>
<td>138</td>
<td>3.7</td>
<td>105</td>
<td>23</td>
<td>0.2</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>02/05/87</td>
<td>17</td>
<td>0.5</td>
<td>11</td>
<td>138</td>
<td>4.8</td>
<td>109</td>
<td>22</td>
<td>0.1</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03/04/87</td>
<td>44</td>
<td>0.6</td>
<td>5 L</td>
<td>140</td>
<td>4.1</td>
<td>106</td>
<td>23</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>103</td>
<td>B</td>
<td>01/30/87</td>
<td>11</td>
<td>0.5</td>
<td>13</td>
<td>140</td>
<td>4.1</td>
<td>103</td>
<td>25</td>
<td>0.1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>02/13/87</td>
<td>25</td>
<td>0.6</td>
<td>7</td>
<td>139</td>
<td>3.9</td>
<td>105</td>
<td>22</td>
<td>0.1</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>02/20/87</td>
<td>32</td>
<td>0.6</td>
<td>10</td>
<td>138</td>
<td>3.7</td>
<td>105</td>
<td>24</td>
<td>0.2</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>03/19/87</td>
<td>59</td>
<td>0.5</td>
<td>9</td>
<td>135</td>
<td>4.3</td>
<td>102</td>
<td>25</td>
<td>0.2</td>
<td>18</td>
</tr>
</tbody>
</table>
**FIGURE 2**

**SAMPLE OUTPUT FROM SUMSTAT MACRO**

**SUMMARY STATISTICS**

Center = 4

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Summary Statistics</th>
<th>Drug A</th>
<th>Drug B</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>N</td>
<td>42</td>
<td>42</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>27.29</td>
<td>27.36</td>
<td>27.32</td>
</tr>
<tr>
<td></td>
<td>Std. Dev.</td>
<td>2.79</td>
<td>2.83</td>
<td>2.80</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>26.91</td>
<td>26.89</td>
<td>26.89</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>22.87</td>
<td>23.02</td>
<td>22.87</td>
</tr>
<tr>
<td></td>
<td>Maximum</td>
<td>33.15</td>
<td>36.49</td>
<td>36.49</td>
</tr>
<tr>
<td>Age Category (years)</td>
<td>21-25</td>
<td>14 (33.3%)</td>
<td>16 (38.1%)</td>
<td>30 (35.7%)</td>
</tr>
<tr>
<td></td>
<td>26-30</td>
<td>23 (54.8%)</td>
<td>22 (52.4%)</td>
<td>45 (53.6%)</td>
</tr>
<tr>
<td></td>
<td>31-35</td>
<td>5 (11.9%)</td>
<td>3 (7.1%)</td>
<td>8 (9.5%)</td>
</tr>
<tr>
<td></td>
<td>36-40</td>
<td>1 (2.4%)</td>
<td>1 (2.4%)</td>
<td>1 (1.2%)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>24 (57.1%)</td>
<td>25 (59.5%)</td>
<td>49 (58.3%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>18 (42.9%)</td>
<td>17 (40.5%)</td>
<td>35 (41.7%)</td>
</tr>
<tr>
<td>Race</td>
<td>White</td>
<td>33 (78.6%)</td>
<td>35 (83.3%)</td>
<td>68 (81.0%)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>4 (9.5%)</td>
<td>3 (7.1%)</td>
<td>7 (8.3%)</td>
</tr>
<tr>
<td></td>
<td>Hispanic</td>
<td>4 (9.5%)</td>
<td>4 (9.5%)</td>
<td>8 (9.5%)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>1 (2.4%)</td>
<td></td>
<td>1 (1.2%)</td>
</tr>
</tbody>
</table>

**FIGURE 3**

Sample VIBLES file for SUMSTAT macro. The VIBLES data set supplies information on formats, labels, etc., to the macro.

data vbles ;
input @1 type 1. vble $8. label $30. format $10. nc $2. ;
cards ;
0studynum Center centerf.
1treatmnt Treatment trtfmt.
2age Age (years) n
2agecat Age Category (years) agefmt. c
2sex Sex sexfmt. c
2race Race racefmt. c