ABSTRACT

SAS programmers in the pharmaceutical industry are frequently under time constraints in generating statistical tables for clinical reports. Most tables are similar in format. If we create a generic table generation program, we can greatly reduce the statistical table generating time. This paper discusses how to create generic statistical tables generating programs for clinical reports by using SAS macro techniques and the REPORT procedure (PROC REPORT). There are two major macro structures used in the provided example, a demographic table program. Each macro determines descriptive and/or inferential statistics for continuous variables and categorical variables. The statistical analysis for the categorical variables includes sample sizes, percentages and probability values (p-value) generated by the FREQ procedure, while the analysis for the continuous variables includes means, standard deviations (STD), ranges and p-values generated by the MEANS procedure and the GLM procedure. This example shows a generic way to build demographic tables, and the method can be applied to other tables that display continuous variables and/or categorical variables.

INTRODUCTION

Many programmers are revising old programs to fit new requests. Some of the programs are easy to revise, and some could lead one into a maze of disaster. A well structured program can make the revision smooth, easy, and time saving. This paper introduces a method to compose a program in a generic structure for report building. There are four steps in this table generation method:

- Prepare the data: Read in data from one or more data sets and create new variables when needed.
- Write the macros which perform the analysis requested.
- Make proper macro calls.
- Output the table by PROC REPORT.

As an example, demographic table building is introduced. A demographic table is requested and its template is as in Chart I.

Chart I:

Table 1
Patient Demographics: Age, Sex, Race, and Height

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study Drug</th>
<th>Vehicle</th>
<th>P-value[a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Years)</td>
<td>N</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>XX.X</td>
<td>XX.X</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>X.XX</td>
<td>X.XX</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td>Sex</td>
<td>Male</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td>Race</td>
<td>Caucasian</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td></td>
<td>Black</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td></td>
<td>Caucasian</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td></td>
<td>Non-caucasian</td>
<td>XX(XX.X%)</td>
<td>XX(XX.X%)</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>N</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>XX.X</td>
<td>XX.X</td>
</tr>
<tr>
<td></td>
<td>STD</td>
<td>X.XX</td>
<td>X.XX</td>
</tr>
<tr>
<td></td>
<td>Min</td>
<td>XX</td>
<td>XX</td>
</tr>
<tr>
<td></td>
<td>Max</td>
<td>XX</td>
<td>XX</td>
</tr>
</tbody>
</table>

[a] P-values for between-treatment comparisons. Age and height were analyzed by general linear models. Sex and race (Caucasian vs. Non-caucasian) were analyzed by chi-square.
DATA PREPARATION

There are four variables to be analyzed and presented in the request: age, sex, race, and height. Two of them are categorical (sex and race) and two are continuous (age and height). If the four variables are in the data sets read in, only one additional race variable (Caucasian vs. Non-Caucasian) needs to be created to calculate the p-value. In the example, a dummy data set is created by using the SAS/Basic DATA step CARD statement. For programming codes, see Appendix A, Part 1.

MACROS

There are two major macro pieces to be created for statistical analysis:

1. For continuous variables:
   A variable analyzed to show its mean score is considered a continuous variable in this topic. Use PROC MEANS to output the descriptive statistics into a SAS data set, then transpose the data set into a treatment-column structure by PROC TRANSPOSE. Retrieve inferential-statistics' p-values from the statistics data set output by PROC GLM. Then, merge the two statistical data sets and store the merged data set into a new data set (named "NEW") by PROC APPEND.

2. For categorical variables:
   A variable analyzed to show its response-frequency distribution is considered a categorical variable in this topic. Use PROC FREQ with option CHISQ to output descriptive and inferential statistics data sets. The output statistics include sample size, percentage, and p-values. Then, transpose the counts and percentages into a treatment-column structure by PROC TRANSPOSE. Perform the necessary manipulation on the p-values, such as rounding the decimal digits, converting its type into character, and assigning a value of “N/A” if the p-value is not calculated, etc. Finally merge the counts and percentages with the p-values and store the merged data set into the "NEW" data set by PROC APPEND.

Since methods to calculate p-values vary, different macro subsets for alternate p-values, such as CHISQ, CMH, Fisher's 2x2 EXACT, Anova, Wilcoxon, etc., should be prepared for use. Only local variables should be used in these macro pieces to support report building generically. For programming codes, see Appendix A, Part 2.

PROPER MACRO CALLS

Make macro calls in the same sequence to be presented on the table. Macro calls are made in the example as follows:

%cont(age,Age (Years), 1);
%catg(sex,Sex,sex.,chisq,2);
%catg(race,Race,racea,chisq,3);
%catg(race2,Race,racea,chisq,3);
%cont(ht,Height,4);

After each macro call finishes running, the output records are stored in the "NEW' data set create by PROC APPEND. When all the macro calls finish running, the records in "NEW' data set are as listed in chart II.

OUTPUT TABLE

PROC REPORT is employed to format the table in the way requested. The table displayed in chart III is generated by the example program running under a VAX/VMS environment and by SAS System version 6.08. For programming codes, see Appendix A, Part 4.

<table>
<thead>
<tr>
<th>OBS</th>
<th>1</th>
<th>2</th>
<th>CAT</th>
<th>ORDER</th>
<th>VARIAB</th>
<th>CATN</th>
<th>PVALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
<td>10</td>
<td>N</td>
<td>1</td>
<td>Age</td>
<td>1</td>
<td>0.678</td>
</tr>
<tr>
<td>2</td>
<td>40.3</td>
<td>46.7</td>
<td>Mean</td>
<td>1</td>
<td>Age</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>3</td>
<td>8.26</td>
<td>9.94</td>
<td>STD</td>
<td>1</td>
<td>Age</td>
<td>3</td>
<td>0.613</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>32</td>
<td>Min</td>
<td>1</td>
<td>Age</td>
<td>4</td>
<td>0.613</td>
</tr>
<tr>
<td>5</td>
<td>51</td>
<td>62</td>
<td>Max</td>
<td>1</td>
<td>Age</td>
<td>5</td>
<td>0.613</td>
</tr>
<tr>
<td>6</td>
<td>5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>Male</td>
<td>2</td>
<td>Sex</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>7</td>
<td>5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>Female</td>
<td>2</td>
<td>Sex</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>8</td>
<td>5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>Caucasian</td>
<td>3</td>
<td>Race</td>
<td>1</td>
<td>1.000</td>
</tr>
<tr>
<td>9</td>
<td>4 (40.0%)</td>
<td>4 (40.0%)</td>
<td>Black</td>
<td>3</td>
<td>Race</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>10</td>
<td>1 (10.0%)</td>
<td>1 (10.0%)</td>
<td>Other</td>
<td>3</td>
<td>Race</td>
<td>3</td>
<td>1.000</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>1</td>
<td>N</td>
<td>1</td>
<td>Race</td>
<td>4</td>
<td>1.000</td>
</tr>
<tr>
<td>12</td>
<td>5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>Caucasian</td>
<td>3</td>
<td>Race</td>
<td>5</td>
<td>1.000</td>
</tr>
<tr>
<td>13</td>
<td>5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>Non-caucasian</td>
<td>3</td>
<td>Race</td>
<td>6</td>
<td>1.000</td>
</tr>
<tr>
<td>14</td>
<td>10</td>
<td>10</td>
<td>N</td>
<td>4</td>
<td>Height</td>
<td>1</td>
<td>0.637</td>
</tr>
<tr>
<td>15</td>
<td>171.7</td>
<td>168.4</td>
<td>Mean</td>
<td>4</td>
<td>Height</td>
<td>2</td>
<td>1.000</td>
</tr>
<tr>
<td>16</td>
<td>8.30</td>
<td>7.46</td>
<td>STD</td>
<td>4</td>
<td>Height</td>
<td>3</td>
<td>1.000</td>
</tr>
<tr>
<td>17</td>
<td>158</td>
<td>158</td>
<td>Min</td>
<td>4</td>
<td>Height</td>
<td>4</td>
<td>1.000</td>
</tr>
<tr>
<td>18</td>
<td>185</td>
<td>182</td>
<td>Max</td>
<td>4</td>
<td>Height</td>
<td>5</td>
<td>1.000</td>
</tr>
</tbody>
</table>
Table 1

Patient Demographics: Age, Sex, Race, and Height

<table>
<thead>
<tr>
<th>Variable</th>
<th>Study Drug</th>
<th>Vehicle</th>
<th>P-value[a]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (Years)</td>
<td>N 10</td>
<td>10</td>
<td>0.678</td>
</tr>
<tr>
<td></td>
<td>Mean 40.3</td>
<td>46.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 8.26</td>
<td>9.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min 29</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max 51</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>Male 5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Female 5 (50.0%)</td>
<td>5 (50.0%)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td>Caucasian 5 (50.0%)</td>
<td>5 (50.0%)</td>
<td>1.000</td>
</tr>
<tr>
<td></td>
<td>Black 4 (40.0%)</td>
<td>4 (40.0%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other 1 (10.0%)</td>
<td>1 (10.0%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Caucasian 5 (50.0%)</td>
<td>5 (50.0%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Non-caucasian 5 (50.0%)</td>
<td>5 (50.0%)</td>
<td></td>
</tr>
<tr>
<td>Height (cm)</td>
<td>N 10</td>
<td>10</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>Mean 171.7</td>
<td>168.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>STD 8.30</td>
<td>7.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Min 158</td>
<td>158</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Max 185</td>
<td>182</td>
<td></td>
</tr>
</tbody>
</table>

[a] P-value for between-treatment comparisons. Age and height were analyzed by general linear model. Sex and race (Caucasian vs. Non-caucasian) were analyzed by chi-square.

CONCLUSION

The programs, structured into the four steps described above, have clear logic flow and are easy to adopt from one project to another. Revising existed programs becomes smooth, easy, and time saving by changing the input data, macro calls and the report procedure (titles and footnotes, etc.). The macro pieces can be reused by other programs. Since each table has its own characteristics, it is necessary to modify macro pieces to fit the table. It would be convenient to set up a macro library storing a collection of different purpose macro pieces for use. This method can be used for many of the tables in an NDA submission and other similar reports.

APPENDIX A

options ls=88 ps=49 nodate nonumber mprint;
proc format;
  value sex 1='Male' 2='Female';
  value race 1='Caucasian' 2='Black' 3='Other';
  value racea 1='Caucasian' 2='Non-caucasian';
run;
* Part 1: Get demographic data ready;
data demo;
  input ptid $ site trt sex race age ht;
  if race=1 then race2=1;   * Caucasion;
  if race<>1 then race2=2;   * Non-caus.;
cards;
  proc sort; by trt; run;
* Part 2: Macro pieces for calculating statistics;
%macro cont(invar,label,order);
  proc printto print='"g&order.. dat"';
  proc means data=demo noprint;
  run;
%mend cont;
options ls=88 ps=49 nodate nonumber mprint;
by trt;
var &invar;
output out=&invar n=nn mean=nmean
std=nstd min=nmin max=nmax;
proc printto; run;
data &invar(drop=nn nmean nstd nmin nmax);
    length n mean std min max $12 varlab $12;
    set &invar;
    order=&order;
    varlab="&label";
    mean=put(nmean,5.1);
    std=put(nstd,6.2);
    array v{3} nn nmin nmax;
    array c{3} $ n min max;
    do i=1 to 3;
        c{i}=put(round(v{i},.1),3.);
    end;
run;
proc transpose data=&invar out=&invar.t;
    id trt;
    by order varlab;
    var n mean std min max;
run;
data &invar.t(drop=a b _name_);
    length _1 _2 $12 cat $13;
    set &invar.t(rename=(_l=a _2=b));
    _1=a;
    _2=b;
    select (_name_);
        when ('N') do; cat='N'; catn=1; end;
        when ('MEAN') do; cat='Mean'; catn=2;end;
        when ('STD') do; cat='STD'; catn=3; end;
        when ('MIN') do; cat='Min'; catn=4; end;
        when ('MAX') do; cat='Max'; catn=5; end;
        otherwise;
    end;
run;
proc sort data=&invar.t; by order catn; run;
data &invar;
    merge &invar.t &invar.p;
    by order;
    if not first.order then pvalue='';
    run;
proc append base=new data=&invar force; run;
%macro catg(invar,label,fmt,method,order);
    proc printto print="g&order..dat";
    proc freq data=demo noprint;
        tables trtiout=trt;
        tables trt*&invar/out=&invar;
    proc printto; run;
data _null_; run;
%mend cont;
proc printto print="g&order..dat";
proc freq data=demo noprint;
tables trt/out=trt;
tables trt*&invar/out=&invar;
proc printto; run;
data &invar;
    set &invar;
    order=&order;
run;
data &invar;
    set &invar;
    order=&order;
run;
proc sort data=&invar; by order trt; run;
data &invar;
    proc glm data=demo outstat=&invar.p;
        model &invar=trtlsite/ss3;
    proc printto; run;
data &invar.p(keep=order pvalue);
    length pvalue $8;
    set &invar.p;
    if _type_='SS3' and _source_='TRT';
    order=&order;
    pvalue=put(round(prob,.001),8.3);
run;
data &invar;
    merge &invar.t &invar.p;
%if upcase(&method)=CHISQ %then %do;
    proc printto print='g&order.. dat';
    proc freq data=demo noprint;
        tables trt*&invar/chisq;
        output out=&invar.p chisq;
    proc printto; run;

    data &invar.p(keep=order pvalue);
        set &invar.p;
        length pvalue $8;
        order=&order;
        pvalue=put(p_pchi,8.3);
    run;

    data &invar(drop=&invar);
        length pvalue $8;
        set &invar.t(drop=_name_);
        by order;
        if not first.order then pvalue=' ';
    run;
%end;
%else %if upcase(&method)= %then %do;
    data &invar(drop=&invar);
        length pvalue $8;
        set &invar.t(drop=_name_);
        by order;
        pvalue=' ';
        output;
        if last.order then do;
            cat=' '; _1=' '; _2=' '; pvalue=' ';
            output;
        end;
    run;
%end;

proc append base=new data=&invar force; run;
%mend catg;

* Part 3: Statistical analysis by macro calls;

%cont(age,Age (Years),1);
%catg(sex,Sex,sex.,chisq,2);
%catg(race,Race,racea.,chisq,3);
%catg(race2,Race,racea.,chisq,3);
%cont(ht,Height,4);

* Part 4: Output table;

proc report data=new headskip;
    column order varlab cat _1 _2 pvalue;
    break after order/skip;
    define order/order noprint;
    define varlab/order width=12 left 'Variable';
    define cat/display spacing=3 width=13 left ' ';
    define _1/display spacing=3 width=12 left 'Study Drug';
    define _2/display spacing=3 width=12 left 'Vehicle';
    define pvalue/display spacing=4 width=10 left 'P-value[a]';
    title 'Table 1';
    title3 'Patient Demographics: Age, Sex, Race, and Height';
    compute after;
        line @1 88 *' ':
        line @1 ' [a] P-value for between-treatment comparisons. Age and height were analyzed by general';
        line @1 ' linear model. Sex and race (Caucasian vs. Non-caucasian) were analyzed by chi-square.';
    endcomp;
run;

ACKNOWLEDGMENTS

The following staff members of Allergan Inc. contributed to the preparation of this paper:

James Wang, Ph.D.,
Ron DeGryse, M.S., M.A..

SAS is a registered trademark or trademark of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other product names are registered trademarks or trademarks of their respective companies.

CONTACT

Irene Zhao
Allergan Inc.
2525 Dupont Dr.
Irvine, CA 92623
Phone (714)246-6572
E-mail: irenez@mail024.usirvine.allergan.sprint.com