%SummaryTable: A SAS® Macro to Produce a Summary Table in Clinical Trial

Yinmei Zhou, St. Jude Children’s Research Hospital, Tennessee
Lijun Zhang, University of Memphis, Memphis, Tennessee
Michael L. Hancock, St. Jude Children’s Research Hospital, Tennessee

ABSTRACT

A summary table with statistical comparisons is one of the most common tables included in clinical study reports. The table usually contains not only summary statistics such as, count, mean, standard deviations, medians, ranges for relevant continuous variables of interest, and number and percentage of patients falling within a particular category for categorical variables of interest, but also some statistical information such as p-values from Chi-square test, Fisher’s exact test, t-test, ANOVA test, etc. We have developed a user-friendly macro program to calculate summary statistics and test results for both continuous and categorical variables of interest, and output a highly customized, presentable table into Microsoft Word. The SAS Macro program is flexible, allowing you to choose the descriptive statistics and the statistical methods for comparisons among groups.

INTRODUCTION

In almost every clinical trial, a summary table with statistics comparisons is always reported before performing any statistical analysis among treatments. In a randomized clinical trial, the summary data and comparison of baseline characteristics are always used to examine how potential differences in baseline characteristics may explain the treatment outcome. A program to automate the generation of such a table is important to help ensure accuracy and reduce programming time.

There are many programs that produce tables from summary statistics\textsuperscript{1,2}, but none of them have the options to print p-values from a specified test. In addition, those programs do not have an easily understood and explicit interface for you to generate the desired output.

The SAS Macro described in this paper uses some basic SAS procedures and the RTF destination to generate a summary table which concisely summarizes key statistics and the tests specified by you from: Chi-square test, Fisher's exact test, Kruskal-Wallis test for categorical variables and 2-sample student's t-test, ANOVA, and Wilcoxon-rank sum test for continuous variables. You have the options to choose the type of test. With an explicit macro interface, you do not need to understand how SAS produces the output, and the only part for you to fill in is the following parameters for macro function.

MACRO SYNTAX

```
%SummaryTable(DATA=, ROW=, RTYPE=, COL=, TEST=, PCT=, RTFFILE=, TITLE=, PAGE=, PGS=)
```

MACRO PARAMETERS

- **DATA =** specifies the name of the macro input data set (required)
- **ROW =** specifies a list of variables that you want to have the statistics summary (required). These variables can be continuous or categorical
- **RTYPE=** specifies the corresponding variables’ type in the row (required) cont - continuous variables; cate - categorical variables
Paper AD13

**FEATURES**

1) **CLEAN INTERFACE**

You need only enter a few parameters to the macro to generate the table. These parameters are straightforward and easy to understand. Even though the macro program consists of several programs, you only need to interact with macro function “%SummaryTable” as described above.

2) **FLEXIBLE AND POWERFUL TO GENERATE OUTPUTS IN A DESIRABLE LAYOUT AND FORMAT**

You can choose the corresponding type of test for each variable of interest to print the p-values. If you do not need the p-values, then input “NA” as the value of parameter “test”. To control how the percentage is calculated, you just need to choose value of parameter “pct” from “col” and “row”. Missing values of the row variables will be included in percentage calculations and printed in a separate row, but be excluded from tests, which will highly help you to control the desirable comparisons.

The outputs layout can be totally decided through the macro function parameters, which include the number of row in each page, page orientation (portrait or landscape), and the output order of variables. The format and label for each column variable and row variable can be controlled through the input data set, and if there are no specified labels, the variable’s name will be used.

**TECHNICAL DETAILS**

**STEP1. READ IN DATA**

All the column variables and row variables will be read in through the macro call (%var1). Each row variable, based on the type of variable (categorical or continuous), will be passed to two different macro functions: %cont_file and %cate_file, which will generate descriptive statistics for each variable and save them in a data set.

```sas
%LET J=1;
%LET ROWJ=%UPCASE(%SCAN(&ROW,&J));
%DO %WHILE(%LENGTH(&ROWJ) > 0);
  %LET RTYPEJ=%UPCASE(%SCAN(&RTYPE,&J));
  %LET TESTJ=%UPCASE(%SCAN(&TEST,&J));
  %VAR1(DATA=&DATA, ROW=&ROWJ, COL=&COL, RTYPE=&RTYPEJ, TEST=&TESTJ, TABLE=_STORE);
  %LET J=%EVAL(&J + 1);
%END;

......

%MACRO VAR1(DATA=, ROW=, RTYPE=, COL=, TEST=, TABLE=_STORE);
  ......%IF &RTYPE=CONT %THEN %DO;
    %CONT_FILE(DATA=&DATA, ROW=&ROW, COL=&COL);
  %END;
%ELSE %IF &RTYPE=CATE %THEN %DO;
    %CATE_FILE(DATA=&DATA, ROW=&ROW, COL=&COL);
  %END;
  ......%MEND;
```

Part of the SAS Macro for input of variables
STEP 2. TEST P-VALUE GENERATION

Through the macro interface, you will specify a test for each row variable. As there are different test types for categorical and continuous variables, the program will check if the test type is compatible with variable type, and if not, an error message will be written in the LOG file and the program will stop running. If there is no error, a test will be performed and the test’s P-value will be saved in a data file for later use in the output table. The P-values are saved in a data set.

An example code for “Chi-square” test

```sas
%IF &FREQTEST=CHISQ %THEN %DO ;
ODS OUTPUT "CHI-SQUARE TESTS"=CHISQ;
PROC FREQ DATA=DATA0 ;
   TABLES &ROW * &COL / NOROW NOCOL NOPERCENT & FREQTEST;
RUN;
ODS OUTPUT CLOSE;
%END ;
%IF %UPCASE(&TEST)=CHISQ %THEN %DO;
DATA _PVALUE(KEEP=PVALUE TEST);
   SET CHISQ;
   IF STATISTIC="CHI-SQUARE" THEN DO;
      PVALUE=PROB;
      TEST='P';
      OUTPUT;
      STOP;
   END;
RUN;
%END;
```

An example code for “Chi-square” test

STEP 3. CONGLOMERATION AND OUTPUT

For each row variable, step 1 and 2 generated basic statistics under each level of column variable, and stored the P-value for the specified test in a data set, respectively. Once the two data sets are merged together, with your required output layout format, a simple, concise and user-friendly summary table in RTF file would be generated.

```sas
PROC REPORT DATA=&DAT NOWD HEADLINE HEADSKIP SPLIT='#' BOX MISSING;
COLS
( "^S={}\BRDRB\BRDRS"
   HED
   TEM
   FOT
   RLABEL
   PRTTOTAL
   ("\B\FS16 &LABELC" %NAMELS2(&CLAB,&NCOL) )
   PVALUE );
DEFINE HED / GROUP NOPRINT ORDER=DATA;
DEFINE TEM / GROUP NOPRINT ORDER=DATA;
DEFINE FOT / GROUP NOPRINT ORDER=DATA;
DEFINE RLABEL / DISPLAY "VARIABLES" STYLE=[JUST=L FONT_WEIGHT = LIGHT];
DEFINE PRTTOTAL / DISPLAY " TOTAL#(N=&T_NUM)”CENTER STYLE(COLUMN)=[JUST=C];
%DO I=1 %TO &NCOL;
   DEFINE &&CLAB&I /DISPLAY "&&CLAB&I.#%STR((N=&&CCNT&I))"STYLE(COLUMN)=[JUST=R];
%END;
DEFINE PVALUE_ / DISPLAY ' P-VALUE*' STYLE(COLUMN)=[ JUST=R] ;
BREAK BEFORE HED /;
COMPUTE BEFORE _PAGE_ /STYLE={(JUST=C FONT_SIZE=8PT )};
   LINE "^S={}\N";
   LINE "\R\B\FS16\CF13 &TITLE";
ENDCOMP;
```
Codes for table generation

EXAMPLE

This Macro function is especially useful for a large data set and huge time saver, but here in order to keep output simple, an example is used for illustration. In this data set, gender, lungcancer, race, and grade are the categorical variables, and they can be numeric (e.g., gender, lungcancer) or character (e.g., race, and grade); age, and height are the continuous variables. We will use the lungcancer (yes and no) as the group (column) variable.

```sas
PROC FORMAT;
   VALUE LUNGFMT
      1="NO"
      2="YES";
   VALUE SEXFMT
      1='FEMALE'
      2='MALE';
RUN;

DATA TEST;
   INPUT GENDER LUNGCANCER AGE HEIGHT GRADE$ @@;
   DATALINES;
      1 1 15 165 L
      1 1 24 140 L
      . 2 30 200 S
      2 . 20 175 L
      2 2 24 153 S
      1 1 10 215 L
      1 2 15 121 S
   ;
   DATA TEST;
   SET TEST;
   FORMAT GENDER SEX. LUNGCANCER LUNGFMT.;
   LABEL AGE='AGE (YR.)'
         SEX='GENDER'
         LUNGCANCER='LUNG CANCER';
RUN;
```
Once the data are prepared, you can apply this macro on the data set to generate a report. In the following macro calling interface, we input the column variable, row variable, type of row variable, and test type to request the summary statistics and the test P-value. The other parameters will control how the output formatted, such as what is the direction of the page (parameter: page), how many rows on each page (parameter: pgs), what is the title for the output table (parameter: title) and where the output will be saved in (parameter: rttfile).

```
%INCLUDE "USERS\MACROS\PROGRAMS\INCLUDE.TABLE.TXT";

%SUMMARYTABLE(
  DATA=TEST,     
  COL=LUNGCA\NCANCER,  
  ROW=GENDER GRADE AGE HEIGHT, 
  RTYPE=CATE CATE CONT CONT, 
  TEST=CHISQ EXACT TTEST ANOVA, 
  PCT=COL,        
  RTFFILE=USERS\MACROS\PROGRAMS\SUMMARYTEST.RTF, 
  TITLE=THE SUMMARY STATISTICS AND COMPARISONS BETWEEN GROUPS, 
  PAGE=PORTRAIT, 
  PGS=20         
);
```

The macro will produce three major groups of statistics:

The total number of patients in each group (column variable).

The summary statistics: the program will report the number and percentage (column percentage or row percentage) of the patients for categorical variable, and report the means, standard deviations, median, the minimum, and the maximum for continuous variable.

The statistical test results - p-values: you can choose a Chi-square test, fisher’s exact test, likelihood ratio chi-square test, and Kruskal-Wallis test for a categorical variable, or choose a t-test, ANOVA, or Wilcoxon Rank-sum test for a continuous variable. The program is very flexible. If you do not want to report the p-values, you can easily enter “NA” as the value for parameter test.

A RTF file of “SummaryTest.rtf” will be created after you run the program, and the output table is in a very nice format commonly seen in medical journals.
The Summary Statistics and comparisons between Groups

<table>
<thead>
<tr>
<th>Variables</th>
<th>Total (n=7)</th>
<th>NO (n=3)</th>
<th>YES (n=4)</th>
<th>P-value*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GENDER [n (%)]</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4 (57.1)</td>
<td>3 (100.0)</td>
<td>1 (25.0)</td>
<td>0.0833 (P)</td>
</tr>
<tr>
<td>Male</td>
<td>2 (28.6)</td>
<td>0 (0)</td>
<td>2 (50.0)</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>1 (14.3)</td>
<td>0 (0)</td>
<td>1 (25.0)</td>
<td></td>
</tr>
<tr>
<td><strong>GRADE [n (%)]</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.1429 (F)</td>
</tr>
<tr>
<td>L</td>
<td>4 (57.1)</td>
<td>3 (100.0)</td>
<td>1 (25.0)</td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>3 (42.9)</td>
<td>0 (0)</td>
<td>3 (75.0)</td>
<td></td>
</tr>
<tr>
<td><strong>AGE (YR.)</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.2969 (T)</td>
</tr>
<tr>
<td>N</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>19.7</td>
<td>16.3</td>
<td>22.3</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>6.8</td>
<td>7.1</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>20</td>
<td>15</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>10.0</td>
<td>10.0</td>
<td>15.0</td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>30.0</td>
<td>24.0</td>
<td>30.0</td>
<td></td>
</tr>
<tr>
<td><strong>HEIGHT</strong></td>
<td></td>
<td></td>
<td></td>
<td>0.6994 (A)</td>
</tr>
<tr>
<td>N</td>
<td>7</td>
<td>3</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>167.0</td>
<td>173.3</td>
<td>162.3</td>
<td></td>
</tr>
<tr>
<td>Std</td>
<td>32.9</td>
<td>38.2</td>
<td>33.5</td>
<td></td>
</tr>
<tr>
<td>Median</td>
<td>165</td>
<td>165</td>
<td>164</td>
<td></td>
</tr>
<tr>
<td>Min</td>
<td>121.0</td>
<td>140.0</td>
<td>121.0</td>
<td></td>
</tr>
<tr>
<td>Max</td>
<td>215.0</td>
<td>215.0</td>
<td>200.0</td>
<td></td>
</tr>
</tbody>
</table>

Program: (test.sas)  Date: (18NOV05)  Page 1 of 1

A Analysis of Variance  F Fisher's Exact Test  P Pearson's Chi-Square Test  T T Test

Example of Macro Output

REFERENCES


Your questions and comments are very much appreciated. We can be reached at:

Yinmei Zhou
Department of Biostatistics
St. Jude Children's Research Hospital
332 N. Lauderdale St.
Memphis, TN 38105-2794
901.495.3706(o)
yinmei.zhou@stjude.org