

A Set of SAS® Macros for Producing Statistical Reports

Greg Grandits

Division of Biostatistics, University of Minnesota

ABSTRACT

Monitoring clinical trials requires periodic generation of statistical reports of comparison between treatment groups. These reports usually include summary statistics for each treatment group and an assessment of statistical significance (p-values). Existing SAS procedures such as PROC TABULATE are adequate to produce a report with some summary statistics (counts, means, SD, etc.), but are not adequate for other statistics (relative risks, regression coefficients, etc) or for including p-values. This paper describes a set of macros that produce statistical reports which allows free placement of text and statistical summaries including p-values onto the report page.

The user first defines columns and column widths across the report page. Text or data values (summary statistics) are then moved to these columns and specified lines using macros MOVE and NMOVE. Summary statistics are available from calling a macro which runs a SAS procedure (GLM, REG, PHREG, etc.), outputs statistics to a SAS dataset, and then compresses the statistics into a one observation dataset. Statistics are placed into array type names which can be moved to the report page after a SET statement. Example: %nmove(P1-P8, col=7, line=12L8) moves 8 p-values to the 7th column defined starting on line 12. A description of how to use the macros and examples of report programs are presented.

INTRODUCTION

In clinical trials periodic reports are generated to monitor study progress and to compare treatment groups for relevant variables of interest. These reports can be varied, but usually contain summary statistics of variables (N's, MEAN's, SD's, etc.), and often contain other statistical information such as p-values, Z-statistics, relative risks, and regression coefficients. It is of interest then for the statistical coordinating center responsible for study reporting to be able to produce these reports using computer software so that no transcribing of numbers, typing, or editing of computer output is necessary. Standard SAS software is in general inadequate to meet these needs. However, SAS capabilities in the dataset using PUT statements and ability to output summary statistics from procedures to SAS datasets provides a mechanism from which macros can be developed to meet these needs. This paper describes a set of macros that produce customized statistical reports that are easy to program and modify, and give complete flexibility to placement of text and data values onto the report page.

The description and use of the basic macros for moving text and data values onto the report page have been given in an earlier SUGI paper (1). These are briefly outlined here and followed by a description of the statistical macros that have been developed that avail the user to statistical information from several SAS procedures, which can be moved to the report page using these earlier developed macros. This is followed by examples giving the program statements and the reports generated by those statements. These set of macros provide a comprehensive set of tools that can be used to generate statistical reports in a variety of research environments.

STEPS IN WRITING A REPORT USING MACROS

Report programs are made up of:

1. a REPORT statement that indicates a new report is starting.
2. a COLSET statement that defines columns and column widths across a report page to which text or data values are later moved.
3. text moves using MOVE. Features include centering, underlining, and repeating text.
4. one or more SET statements that reads in statistical information from a SAS dataset. These one observation datasets will have been generated from one of the statistical generating macros described later.
5. one or more NMOVE statements which move the statistical information onto the report page. NMOVE allows the whole vector of values to be moved with a reference to beginning and ending values, e.g. P1-P20 moves 20 p-values.

SYNTAX FOR REPORT MACROS:

1. REPORT is used simple as %REPORT which indicates the start of a new report
2. COLSET is used as follows:

%COLSET (col1size nospaces"x" col2size nospaces ...);

Ex: %COLSET (25 10 2X 10 2X 10);

This statement sets up 4 columns. The first column is 25 positions long and the last 3 columns are 10 positions long. Two spaces are placed between the last 3 columns. Nothing is written to these positions, and is used to set off text from other text.

3. MOVE is used to place text on the report page and is used as follows:

%MOVE ('string1':'string2':...'stringk', line=, col=, center=, under=, pat=, mfirst=)

This is illustrated by example. More detailed descriptions are given in (1).

'Men': 'Women': 'Total'

text strings to be placed on report page

line = 12 21 33

moves strings to lines 12, 21, and 33

line = 12L3

moves strings to lines 12, 13, and 14

line = 12L6, pat=3rd/1

moves strings to line 12, 13, 14, 16, 17, and 18

col = 3 4 8

moves strings to columns 3, 4, and 8

col = 3 4-8

moves strings to columns 3 and a column formed by combining columns 4-8

col = 2.8

moves strings to columns 2 *through* 8

center set to "Y" for centering text; "N" otherwise

under set to "Y" for underlining text; "N" otherwise

mfirst sets whether you want columns or lines advancing first when strings are moved

Example of %MOVE statement:

```
%MOVE ('Total Cholesterol':LDL-C':HDL-C', col=1, line=10L6, center=n, pat=3rd/1)
```

This statement produces the following text starting at line 10. Note that strings are repeated until the total lines are exhausted.

Total Cholesterol

HDL-C

LDL-C

Total Cholesterol

HDL-C

LDL-C

4. NMOVE places data values on the report page and is used as follows:

```
%NMOVE ( var(1) - var(n), line=, col=, pat=, mfirst=, pat=, fmt=, scaler=);
```

The line=, col=, mfirst=, and pat= statements are identical to those in MOVE. The fmt parameter formats the values; scaler is an optional parameter that multiplies the values by the value given before printing.

Example:

```
set data1;
```

```
%NMOVE ( M1-M50, col = 2.6, line=12L10, fmt=6.2);
```

In this example the dataset variables M1 through M50 are moved to the 5 column by 10 line positions indicated and are printed with the format 6.2.

STATISTIC GENERATING MACROS

Below is a listing of the statistic generating macros available along with the SAS procedure that is called with the macro, the statistics that are available from the output datasets generated, and a brief description of the macro.

Macro	SAS PROC	Statistics Produced	Description of Macro
breakdn	summary	n,mean,sdev, sum,min,max	summary statistics by level of 1 or more class variables
freqdis	summary	counts, percents, cumulative counts, cumulative percents	distribution of variable by level of another variable
glmp	glm	anova p-values, LS means	statistics from analysis of variance
regp	reg	p-values, estimates SE, t-statistics	statistics from linear regression
phregp	phreg	estimates, SE, z-stat's p-values, RRs, 95% CI	statistics from proportional hazards regression
logistp	logist	estimates,SEs, Z-stat's p-values, RRs, 95% CI	statistics from logistic regression
chisqp	freq	Cochran-Mantel-Haenszel p-values	stratified contingency table analysis

Details on use of BREAKDN, GLMP, and PHREGP are given below. Other macros have similar syntax.

%BREAKDN (data=, class=, var=, out=, sfirst=);

This macro reads the SAS dataset specified in DATA using PROC SUMMARY and computes summary statistics for each variable specified in VAR by each level of the variable specified in CLASS. A one observation dataset containing these statistics is written to dataset specified in OUT.

The statistics calculated are N, MEAN, SDEV, SE, SUM, MIN, and MAX. They are contained in the variables N1-N?, M1-M?, S1-S?, SE1-SE?, SUM1-SUM?, MIN1-MIN?, and MAX1-MAX?, where ? depends on the number of variables in VAR and the number of levels in the variables in CLASS.

The parameters specified are illustrated by example.

<code>class = sex 2T</code>	statistics are stored for both levels of the variable SEX and the total
<code>class = sex 2T group 6</code>	statistics are stored for each level of SEX crossed with GROUP. Statistics for each level of group for men and women combined will also be computed.
<code>var = age dbp sbp chol</code>	This is the list of variables to compute statistics for.

DATA is the SAS dataset to be read; OUT is the SAS dataset statistics are written to and contains one observation, and SFIRST indicates the order in which the statistics are stored.

Example: `%BREAKDN (class = sex 2T, var = age dbp sbp chol, out = table1, sfirst=VAR)`

The n's for the 4 variables at sex = 1 are stored in n1-n4.

The n's for the 4 variables at sex = 2 are stored in n5-n8.

The n's for the 4 variables for women and women combined are stored in n9-n12.

The variables are similarly stored for the other statistics.

% GLMP (parameters)

GLMP runs PROC GLM to obtain p-values for factors of interest from the ANOVA table and places them in SAS datasets. The p-values are contained in p1-p? where ? is the number of dependent variables. Least square means are also available and are stored in variables ls1-ls? where ? is the product of the number of levels of the factor of interest and the number of dependent variables.

Parameter	Description
<code>data =</code>	SAS dataset to be read
<code>class =</code>	Class variables for GLM
<code>dlist =</code>	Dependent variable list. An analysis is done for each variable listed.
<code>ilist =</code>	Independent variable list. Used for each dependent variable given in dlist.
<code>blist =</code>	Optional list of variables where, if given, the i'th variable is added to the independent variable list for the i'th analysis.
<code>factor=</code>	Independent variable or variables for which p-values from the ANOVAs are desired.
<code>pdata =</code>	SAS dataset to which p-values are written.
<code>lsmean =</code>	If set to YES, LS means are computed
<code>lsclass =</code>	The class variable for which LS means are computed
<code>lsdata =</code>	SAS dataset for which LS means are written

Example:

`%glmp ( class = group, dlist = tchol hdl ldl trig, ilist = age group, factor=group, pdata=p);`

This runs a PROC GLM for the 4 dependent variables in DLIST with AGE and GROUP as independent variables. P-values for GROUP are output to dataset P which contains one observation with the variables P1-P4.

%PHREGP (parameters)

PHREGP runs PROC PHREG to perform proportional hazards regression and saves results for factors of interest into SAS datasets.

Parameter	Description
data =	SAS dataset to be read
dlist =	Dependent variable list. An analysis is done for each variable listed. The variables are event indicators coded as 1 if event, 0 if censored.
ilist =	Independent variable list. Used for each dependent variable given in dlist.
tlist =	Failure or censoring time list corresponding to events in dlist.
factor=	Independent variable or variables for which statistics from PROC PHREG are to be obtained.
units =	Value regression coefficients are multiplied by before relative risks are computed.
strata =	Optional list of strata variables.
out =	SAS dataset to which statistics are written.

The statistics and the names of the variables that contain them are as follows:

e1-e?	regression coefficients for factor
se1-se?	standard errors of coefficients
z1-z?	z-statistics for factor
p1-p?	p-values for factor
rr1-rr?	relative risks for factor
ul-u?	upper 95% CI for RR
ll-l?	lower 95% CI for RR

where ? is the number of dependent variables

Two complete examples of programs and reports generated by the programs are given at the end of the paper.

Discussion

Much effort has been made by SAS institute and others to make reporting easier. However, there still seems to be a lack of tools to do the statistical reporting needed. The statistical report macros described here are simple to use and have tremendous flexibility. Programs are easy to understand and modify and are typically less than one page (see examples). These macros can also be expanded to include statistics from other procedures since most SAS procedures can output statistics to datasets. The key to making the numeric moves is to get the statistics into a one observation dataset. Then a simple SET statement makes available the information needed. Information from several different sources can easily be included on the report by multiple SET/NMOVE sequences. This gives the flexibility to the system. These macros have proved valuable to clinical trials monitored by the Division of Biostatistics at the University of Minnesota. It is the authors desire that SAS institute consider some of the ideas in these macros and incorporate them in their software.

References:

1. G. Grandits, K. Svendsen. A set of SAS macros for producing customized reports. Proceedings of the 19th Annual SAS Users Group International Conference. 1994;832-837.

*** Assume have dataset called ALL that contains variables needed ***

```
%let listd = age bmb1 sbp1 dbp1 chol2s trig2s hdl2s ldl2s gluc2s;
%let listi = group clinic ;

%breakdn ( data=all, class = group 6T sex 2, var=&listd, out=blmean );
%breakdn ( data=all, class = sex 2 group 6T, var=age , out=bln );

%glmp ( data=all, class = &listi, dlist=&listd, factor=group,
        ilist= &listi, pdata= menp, where = where sex = 1);
%glmp ( data=all, class = &listi, dlist=&listd, factor=group,
        ilist= &listi, pdata= womenp, where = where sex = 2);

%report;
%colset(15 2X 7 7 7 7 7 7 2X 7 7 2X 9 );

%move('Selected Baseline Characteristics of TOMHS':
      'Participants by Treatment Group and Gender', col=1-0, line=3L2);
%move('Treatment Group': 'All Groups', col= 2-7 8-9, line=6, u=yes);
%move('P-value', col=10, line=7);
%move('Ace': 'Aml': 'Chl': 'Dox': 'Enal': 'Plac': 'Mean': 'SD': ' (5 df)',
      col=2.0, line=8, u=yes);
%move('Men': 'Women', col=1, line=8 23, u=yes, center=no);
%move('Age': 'BMI': 'SBP': 'DBP': 'Cholesterol': 'Triglycerides':
      'HDL-C': 'LDL-C': 'Glucose',
      col=1, center=2, line=10L9 25L9);
%move('N', col=1, center=no, line=20 35);

set blmean ;
%nmv(m1-m126 , col=2.8, line=10L9 25L9, fmt=5.1);
%nmv(s109-s126 , col=9 , line=10L9 25L9, fmt=5.1);

set bln;
%nmv(n1-n14, col=2.7 8-9, line=20 35, mfirst=C, fmt=5.0);

set menp;
%nmv(p1-p9, col=10, line=10L9, fmt=7.3);

set womenp;
%nmv(p1-p9, col=10, line=25L9, fmt=7.3);
run;
```

Selected Baseline Characteristics of TOMHS
Participants by Treatment Group and Gender

	Treatment Group						All Groups		P-value (5 df)
	Ace	Aml	Chl	Dox	Enal	Plac	Mean	SD	
Men									
Age	54.8	54.3	55.4	54.1	55.9	55.0	54.9	6.5	0.550
BMI	29.1	29.0	29.1	28.8	28.8	29.2	29.0	3.4	0.945
SBP	140.0	138.2	139.2	140.1	138.5	140.8	139.6	12.1	0.694
DBP	90.8	91.0	90.7	90.7	90.3	90.8	90.8	3.4	0.818
Cholesterol	229.2	228.5	225.7	225.2	221.2	221.2	224.8	36.2	0.538
Triglycerides	144.6	137.4	147.4	140.9	149.9	141.4	143.4	75.1	0.917
HDL-C	39.6	39.7	39.6	40.8	38.8	38.6	39.4	9.0	0.615
LDL-C	160.7	161.2	156.6	156.3	152.4	154.3	156.7	33.5	0.487
Glucose	102.4	100.2	103.8	100.3	103.3	102.6	102.2	16.0	0
N	91	77	87	81	77	144	557		
Women									
Age	54.3	53.0	54.8	55.9	54.4	54.7	54.5	6.3	0.294
BMI	27.6	28.9	28.7	28.4	29.0	28.8	28.6	3.8	0.382
SBP	140.6	138.0	142.7	142.0	143.9	141.7	141.5	12.9	0.206
DBP	90.3	90.7	89.7	90.3	90.0	90.1	90.2	3.4	0.767
Cholesterol	241.8	230.0	236.8	234.2	231.8	231.8	233.8	41.2	0.775
Triglycerides	103.9	114.7	104.5	114.7	129.9	108.7	112.9	75.1	0.445
HDL-C	53.4	48.9	50.4	50.7	50.7	49.5	50.4	12.2	0.598
LDL-C	167.7	158.2	165.4	160.7	155.2	160.6	160.9	36.4	0.561
Glucose	98.6	95.8	96.8	101.4	101.1	99.1	98.9	14.9	0.344
N	41	54	49	53	58	90	345		

*** Assume have dataset called TEMP that contains variables needed ***

```
%let end = chddeath mi angina allcause mrfchd ;
%let tevnt= foptime mrftime angtime foptime mrftime ;
```

```
%phregp(data=temp,dlist=&end,tlist=&tevnt, where = where ecgabn = 1,
        ilist=group,factor=group, out=out1);
```

```
%phregp(data=temp,dlist=&end,tlist=&tevnt, where = where ecgabn = 0,
        ilist=group,factor=group, out=out2);
```

** This call does analyses for interaction **

```
%phregp(data=temp,dlist=&end,tlist=&tevnt,
        ilist=group ecgabn ecg_grp,factor=ecg_grp, out=out3);
```

```
%report;
%colset(20 16 7 7 7 7 3x 12) ;
```

```
%move('Table 2', col=1-0, line=1);
%move('Relative Risk of CHD (Treatment vs Control) by Presence or':
      'Absence of ECG Abnormalities', col=1-0, line=3L2);
%move('Treatment versus Control', col=3-6, line=7, u=1);
%move('Event': 'ECG Status', col=1 2, line=9, u=1, center=2);
%move('RR': 'L95%': 'U95%': 'p-val': 'p-val(int)', col=3.0, line=9, u=1);
%move('CHD Death': 'Non-fatal MI': 'Angina': 'All Mortality':
      'CHD Death or MI', col=1, center=2, line=12L5, pat=1st/2);
%move('Abnormality': 'No Abnormality', col=2,
      line=12L2 15L2 18L2 21L2 24L2, center=2);
```

```
set out1 ;
%nmv(r1-r5 l1-l5 u1-u5, col=3.5, fmt=4.2, line=12L5, pat=1st/2);
%nmv(p1-p5 , col=6, fmt=5.3);
```

```
set out2 ;
%nmv(r1-r5 l1-l5 u1-u5, col=3.5, fmt=4.2, line=13L5, pat=1st/2);
%nmv(p1-p5 , col=6, fmt=5.3);
```

```
set out3 ;
%nmv(p1-p5 , col=7, fmt=5.3);
```

Table 2

Relative Risk of CHD (Treatment vs Control) by Presence or
Absence of ECG Abnormalities

Event	ECG Status	Treatment versus Control				p-val(int)
		RR	L95%	U95%	p-val	
CHD Death	Abnormality	0.45	0.26	0.78	0.004	0.002
	No Abnormality	1.21	0.88	1.66	0.237	
Non-fatal MI	Abnormality	0.88	0.57	1.35	0.562	0.832
	No Abnormality	0.93	0.78	1.11	0.409	
Angina	Abnormality	0.83	0.64	1.08	0.166	0.521
	No Abnormality	0.76	0.68	0.86	0.000	
All Mortality	Abnormality	0.69	0.45	1.05	0.082	0.047
	No Abnormality	1.10	0.90	1.34	0.341	
CHD Death or MI	Abnormality	0.69	0.49	0.97	0.034	0.067
	No Abnormality	0.98	0.84	1.15	0.820	