

Paper 057-2007

Saving Trees with Output Delivery System (ODS)

Angelina D. Tan, Nancy Dilehl, Jay N. Mandrekar
Division of Biostatistics, Mayo Clinic, Rochester, MN

ABSTRACT

Clinical or epidemiological studies often include large number of variables. The analysis may generate several pages of output depending on which SAS[®] procedure is used for the analysis. The SAS[®] Output Delivery System (ODS) is an excellent tool used for formatting output generated by SAS[®] procedures and DATA steps. In this article we highlight delivery of information to the user through creative and innovative techniques using SAS[®] ODS system and provide sample SAS[®] code and macros that are easy to implement and performs the necessary computations without any additional computer time. These concepts can be improvised based on individual project needs. Using data from clinical studies, we illustrate use in the context of regression based procedures that include REG, LOGISTIC and PHREG from SAS[®] and generate reports using procedures. We also generate some simple reports using REPORT procedure for better presentation. Intended audience is expected to have intermediate SAS[®] programming skills.

INTRODUCTION

Analysis involving data from clinical or epidemiological studies, may involve large number of variables. Biostatistician goes through several exploratory analyses before arriving at the final conclusions from the study. Majority of the analysis gets done using SAS[®] as commercial package which produces several pages of output that needs to be reviewed before proceeding to the next step. In a world where business decisions must be made quickly, an application must be able to deliver information in a blink of an eye. Trends, facts and figures must be translated into meaningful information so the meaning can get inside the heads of people. It is inefficient to go through several rows of numbers in an attempt to reveal the hidden meaning of data. Thus data must be provided in a succinct table which is the motivation behind this work.

The SAS[®] Output Delivery System (ODS) is an excellent tool used for generating, storing, reproducing and formatting output generated by SAS[®] procedures and DATA steps. ODS enables effective application design strategies to be implemented while providing flexibility and ease of use. This can be combined with products such as SAS[®]/BASE and SAS[®]/STAT to name a few. Investigators can obtain key summary measures to start out before diving into the details of comprehensive analysis. In this article we highlight delivery of information to the user through creative and innovative techniques using SAS[®] ODS system and provide sample SAS[®] code and macros that are easy to implement and do all the necessary computations without any additional computer time. These concepts can be improvised based on individual project needs. In this paper we illustrate SAS[®] code and macros in the context of REG, LOGISTIC and PHREG procedures from SAS[®] using data from clinical studies and generate reports using procedures such as REPORT for better presentation.

We first provide a brief overview of each of the procedure, data description and present the relevant part of the code and the output that it generates using our SAS[®] code and macro. For technical details on these procedures and the options that they offer we refer reader to the appropriate SAS[®] online documentation. We would like to point out that certain procedures such as MEANS procedure from SAS[®] produce concise output for basic descriptive statistics of interest and thus ODS may not be required for that particular scenario. However, other procedures discussed later in the next sections produces several pages of output if we have several variables and thus we illustrate the way to extract limited information of interest with sample clinical studies.

SOME INSIGHTS ON ODS

ODS statements can be used anywhere in SAS[®] program to control different features of the Output Delivery System. There are three categories of ODS global statement: 1) Output Control; 2) SAS Formatted and 3) Third-party formatted. The output control provides the descriptive information about the specified output objects and their locations, for example, ods select, ods exclude and ods trace. The SAS[®] formatted produces a listing output, a SAS[®] output data set or a hierarchy file, for example, ods listing and ods output. The third-party formatted produces files that are formatted in the proper destination format, for example, ods rtf and ods pdf.

In order to select, exclude, or modify an output object, we must first determine its name. The names of the output object for any SAS/STAT procedure can be obtained 1) from the individual procedure chapter/section of the SAS[®] online Help system; 2) from SAS[®] Explorer window after a SAS[®] run; 3) using the ODS TRACE statement in a SAS[®] run to write identifying information to the SAS[®] log for each generated output object. The following example illustrates the trace record for the specific tables created in this REG procedure step:

```
ods trace on;
  proc reg;
    model y=x;
  run;
ods trace off;
```

PARTIAL SAS® LOG OF TRACE OPTION

Output Added:

```
-----
Name:      ParameterEstimates
Label:     Parameter Estimates
Template:  Stat.REG.ParameterEstimates
Path:     Reg.MODEL.Fit.y.ParameterEstimates
```

Individual output object can be referred with either object name or the fully or partially qualified path. Once output objects and paths have been identified, a subset of output objects can be selected for output with either ODS SELECT or ODS EXCLUDE statement. Then, use ODS OUTPUT statement to produce SAS® data sets from the selected output object for further data manipulations and extractions. When the simplify output is finalized, it can be sent to designate ODS destination, for example, ODS RTF to produce a report in WORD Document.

EXAMPLES SPECIFIC TO SOME REGRESSION BASED SAS® PROCEDURES

In the statistical analysis involving regression modeling, the analysis is first done using univariate model building and then the selection of potential variables to be considered for future model building is made at this stage. Thus we feel it is appropriate to look at the statistics that is relevant to this approach instead of going through the detailed output. If the results from certain univariate model are in question then analyst can run that particular model for further investigation. The tables generated using ODS can assist in variable selection for multivariate model building.

PROC REG

The REG procedure is one of many regression procedures in the SAS® System. It is a general-purpose procedure for regression, while other SAS® regression procedures provide more specialized applications. PROC REG provides the following capabilities: multiple MODEL statements, nine model-selection methods, interactive changes both in the model and the data used to fit the model, linear equality restrictions on parameters, tests of linear hypotheses and multivariate hypotheses, collinearity diagnostics, predicted values, residuals, studentized residuals, confidence limits, and influence statistics, correlation or cross product input, requested statistics available for output through output data sets and various plots of interest.

ILLUSTRATION WITH A SAMPLE DATA

The goal of a study where Aerobic fitness (measured by the ability to consume oxygen) is fit to some simple exercise tests, is to develop an equation to predict fitness based on the exercise tests rather than on expensive and cumbersome oxygen consumption measurements. These measurements were made on men involved in a physical fitness course at N.C.State University. The variables are Age (years), Weight (kg), Oxygen intake rate (ml per kg body weight per minute), time to run 1.5 miles (minutes), heart rate while resting, heart rate while running (same time Oxygen rate measured), and maximum heart rate recorded while running. Oxygen intake rate is a continuous outcome variable. We have 6 continuous predictor variables included in this data set. Thus we can use REG procedure to explore the best linear predictor for measuring the aerobic fitness as a first step in model building process. The relevant SAS® code and output from these univariate linear regression models is presented below.

SAS® CODE

```
*****;
** MACRO NAME: pReg **;
** PARAMETERS: **;
** dsn = Input data set to be analyzed **;
** varlist = Analysis variables **;
** numvars = Number of analysis variables that listed in varlist **;
** respvar = Response variable **;
** outdsn = Output data set that contains the summary statistics **;
** (default=table) **;
*****;

%macro pReg (dsn=, varlist=, numvars=, respvar=, outdsn=table);
  %do j=1 %to &numvars.;
    %let var2=%scan(&varlist.,&j.,' ');
```

```

ods select none;
ods output ParameterEstimates=work.pe&j. ;
proc reg data=&dsn.;
  model &respvar. = &var2. ;
run;
ods select all;

data pe&j.;
  set pe&j.;
  keep Variable Estimate StdErr tValue Probt ;

  if Variable="Intercept" then delete;
run;

%if &j.=1 %then %do;
  data &outdsn.;
  set pel;
run;
%end;
%else %do;
  data &outdsn.;
  set &outdsn. pe&j.;
run;
%end;
%end;
%mend pReg;

*****
**                               MACRO CALL                               **
*****

libname tana 'h:\ibm\sasdata';

%pReg (dsn      =tana.fitness,
      Varlist  =Age Weight RunTime RunPulse RestPulse MaxPulse,
      numvars  =6,
      respvar  =Oxygen,
      outdsn   =preg);

*****
**                               PRODUCE SUMMARY TABLE IN MS WORD DOCUMENT                               **
*****

ods rtf file='h:\ibm\preg.doc';
title1 'Simple Linear Regression Model';
proc report data=preg box nowd split=' ';
  column Variable Estimate StdErr tValue Probt ;
  define Variable / display;
  define Estimate / display;
  define StdErr / display;
  define tValue / display;
  define Probt / display;

  compute probt;
    if probt <= 0.05 then call
      define('_c5_', 'style', 'style=[font_weight=bold]');
  endcomp;
run;
ods rtf close;

```

OUTPUT

Univariate Linear Regression Models

Variable	Parameter Estimate	Standard Error	t Value	Pr > t
Age	-0.31136	0.18080	-1.72	0.0957
Weight	-0.10410	0.11719	-0.89	0.3817
RunTime	-3.31056	0.36119	-9.17	<.0001
RunPulse	-0.20680	0.08852	-2.34	0.0266
RestPulse	-0.27921	0.11903	-2.35	0.0260
MaxPulse	-0.13762	0.10488	-1.31	0.1997

PROC LOGISTIC

Binary responses (for example, success and failure) and ordinal responses (for example, normal, mild, and severe) arise in many fields of study. Logistic regression analysis is often used to investigate the relationship between these discrete responses and a set of explanatory variables. Explanatory response variables may include continuous or discrete variables. Several texts that discuss logistic regression are Collett (1991), Agresti (1990), Cox and Snell (1989), and Hosmer and Lemeshow (1989).

ILLUSTRATION WITH A SAMPLE DATA

We will use the low birth weight study data presented in Hosmer and Lemeshow (2000) for the illustration purposes. Low birth weight is defined as birth weight less than 2500 grams, is an outcome that has been of concern to physicians for years. This is due to the fact that infant mortality rates and birth defect rates are higher for low birth weight babies. Data were collected at Baystate Medical Center in Springfield, Massachusetts. Several variables including mother's behavior during pregnancy, race, age etc. were collected as a part of this study. The data included in this analysis has information on 189 births to women seen at the obstetrics clinic. This includes 59 low birth weight babies. The goal of the original analysis was to identify the set of risk factors from the comprehensive list of variables collected in this study. As a part of this illustration we will focus on the univariate analysis which is often a first step prior to arriving at a multivariable model. Results of interest from these univariate logistic regression models should include p-value, Odds ratio estimate and upper and lower 95% Confidence intervals. However, this information is scattered on two pages and this may get cumbersome as number of variables of interest increases. (This data is posted on <http://www-unix.oit.umass.edu/~statdata/statdata/data/icu.dat>). The relevant SAS[®] code and output from these univariate logistic regression models is presented below.

SAS[®] CODE

```
*****;
** MACRO NAME: pLogistic **;
** PARAMETERS: **;
** dsn = Input data set to be analyzed **;
** varlist = Analysis variables **;
** numvars = Number of analysis variables that listed in varlist **;
** respvar = Response variable **;
** outdsn = Output data set that contains the summary statistics **;
** (default=table) **;
*****;

%macro pLogistic (dsn=, varlist=, numvars=, respvar=, outdsn=table);
  %do j=1 %to &numvars.;
    %let var2=%upcase(%scan(&varlist.,&j.,' '));

    ods select none;
    ods output ParameterEstimates=work.pe&j. OddsRatios=work.or&j. ;
    proc logistic data= &dsn. descending;
      model &respvar.= &var2. /link=logit ;
  %end;
%end;
```

```

run;
ods select all;

data pe&j.;
  length Independent $20;
  set pe&j.;
  rename ProbChiSq=Waldp;
  keep Independent Estimate StdErr WaldChiSq ProbChiSq;

  Independent="&var2.";
  if upcase(Variable)="&var2.";
run;

proc sort data=pe&j.;
  by Independent;
run;

data or&j.;
  length Independent $20;
  set or&j.;
  rename OddsRatioEst=OREst;
  keep Independent OddsRatioEst LowerCL UpperCL;

  Independent="&var2.";
  if upcase(Effect)="&var2.";
run;

proc sort data=or&j.;
  by Independent;
run;

data all&j.;
  merge pe&j. or&j.;
  by Independent;
run;

%if &j.=1 %then %do;
  data &outdsn.;
  set all1;
  run;
%end;
%else %do;
  data &outdsn.;
  set &outdsn. all&j;
  run;
%end;
%end;
%mend pLogistic;

*****;
**                               MACRO CALL                               **;
*****;

libname tana 'h:\ibm\sasdata';

%pLogistic (dsn      =tana.icu,
            Varlist  =AGE SEX SER CAN CRN INF CPR SYS HRA PRE TYP FRA
                    PO2 PH PCO BIC CRE,
            Numvars  =17,
            Respvar  =STA,
            outdsn   =plog);

*****;
**                               PRODUCE SUMMARY TABLE IN MS WORD DOCUMENT                               **;
*****;
ods rtf file='h:\ibm\plogistic.doc';
  title1 'Univariate Logistic regression for endpoint status (Lived,
```

```

        Died)';
proc report data=plog box nowd split=' ';
  column Independent Estimate StdErr WaldChiSq Waldp OREst LowerCL
        UpperCL;
  define Estimate / display;
  define StdErr / display;
  define WaldChiSq / display;
  define Waldp / display;
  define OREst / display;
  define LowerCL / display;
  define UpperCL / display;

  compute orest;
    if waldp <0.05 and orest > 1 then do;
      call define('_c5_', 'style', 'style=[background=yellow]');
      call define('_c6_', 'style', 'style=[background=yellow]');
    end;
  endcomp;
run;
ods rtf close;

```

OUTPUT

Univariate Logistic regression for endpoint status (Lived, Died)

Independent	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Odds Ratio Estimate	Lower 95% Confidence Limit for Odds Ratio	Upper 95% Confidence Limit for Odds Ratio
AGE	0.0275	0.0106	6.7963	0.0091	1.028	1.007	1.049
SEX	0.1054	0.3617	0.0848	0.7708	1.111	0.547	2.258
SER	-0.9469	0.3682	6.6144	0.0101	0.388	0.189	0.798
CAN	0	0.5893	0.0000	1.0000	1.000	0.315	3.174
CRN	1.2198	0.5039	5.8605	0.0155	3.386	1.261	9.091
INF	0.9163	0.3617	6.4172	0.0113	2.500	1.230	5.080
CPR	1.6945	0.5885	8.2913	0.0040	5.444	1.718	17.252
SYS	-0.0170	0.00600	8.0440	0.0046	0.983	0.972	0.995
HRA	0.00294	0.00655	0.2015	0.6535	1.003	0.990	1.016
PRE	0.2339	0.4732	0.2444	0.6211	1.264	0.500	3.194
TYP	2.1846	0.7450	8.5999	0.0034	8.887	2.064	38.272
FRA	0	0.6712	0.0000	1.0000	1.000	0.268	3.726
PO2	0.6601	0.5711	1.3360	0.2477	1.935	0.632	5.927
PH	0.6228	0.6289	0.9808	0.3220	1.864	0.543	6.395
PCO	0	0.5893	0.0000	1.0000	1.000	0.315	3.174
BIC	0.7621	0.5790	1.7327	0.1881	2.143	0.689	6.666
CRE	1.4881	0.6596	5.0903	0.0241	4.429	1.216	16.132

PROC PHREG

The analysis of survival data requires special techniques because the data are almost always incomplete, and familiar parametric assumptions may be unjustifiable. Investigators follow subjects until they reach a pre-specified endpoint (for example, death). However, subjects sometimes withdraw from a study, or the study is completed before the endpoint is reached. In these cases, the survival times (also known as failure times) are censored; subjects survived to a certain time beyond which their status is unknown. The non-censored survival times are referred to as event times. Methods for survival analysis must account for both censored and non-censored data. The PHREG procedure performs regression analysis of survival data based on the Cox proportional hazards model. Cox's semi-parametric model is widely used in the analysis of survival data to explain the effect of explanatory variables on survival times.

ILLUSTRATION WITH A SAMPLE DATA

Krall, Uthoff, and Harley (1975) analyzed data from a study on multiple myeloma in which researchers treated 65 patients with alkylating agents. Of those patients, 48 died during the study and 17 survived. In the data set Myeloma, the variable Time represents the survival time in months from diagnosis. The variable VStatus consists of two values, 0 and 1, indicating whether the patient was alive or dead, respectively, at the end of the study. If the value of VStatus is 0, the corresponding value of Time is censored. The variables thought to be related to survival are LogBUN (log(BUN) at diagnosis), HGB (hemoglobin at diagnosis), Platelet (platelets at diagnosis: 0=abnormal, 1=normal), Age (age at diagnosis in years), LogWBC (log(WBC) at diagnosis), Frac (fractures at diagnosis: 0=none, 1=present), LogPBM (log percentage of plasma cells in bone marrow), Protein (proteinuria at diagnosis), and SCalc (serum calcium at diagnosis). Interest lies in identifying important prognostic factors from these nine explanatory variables. The relevant SAS[®] code and output from these univariate Cox proportional hazards regression models is presented below.

SAS[®] CODE

```
*****;
** MACRO NAME: phReg **;
** PARAMETERS: **;
** dsn = Input data set to be analyzed **;
** varlist = Analysis variables **;
** numvars = Number of analysis variables that listed in varlist **;
** respvar = Response variable **;
** censvar = Censoring variable **;
** censval = List of censoring values (separated by blank) **;
** outdsn = Output data set that contains the summary statistics **;
** (default=table) **;
*****;

%macro phReg (dsn=, varlist=, numvars=, respvar=, censvar=, censval=,
             outdsn=table);
  %do j=1 %to &numvars.;
    %let var2=%upcase(%scan(&varlist.,&j.,' '));

    ods select none;
    ods output ParameterEstimates=work.pe&j.;
    proc phreg data = &dsn. ;
      model &respvar.*&censvar.(&censval.)= &var2./rl;
    run;
    ods select all;

    data pe&j.;
      length Independent $15 ;
      set pe&j.;
      keep Estimate StdErr Independent ProbChiSq HazardRatio
            HRLowerCL HRUpperCL;

      Independent="&var2.";
      if upcase(Variable)="&var2.";
    run;

    proc sort data=pe&j.;
      by Independent;
    run;
```

```

data all&j.;
  merge pe&j. ;
  by Independent;
run;

%if &j.=1 %then %do;
  data &outdsn.;
  set all1;
  run;
%end;

%else %do;
  data &outdsn.;
  set &outdsn. all&j.;
  run;
%end;
%end;
%mend phReg;
*****
**                               MACRO CALL                               **
*****

libname tana 'h:\ibm\sasdata';

%phReg (dsn      =tana.Myeloma,
        Varlist  =LogBUN HGB Platelet Age LogWBC Frac LogPBM Protein
                SCalc,
        Numvars  =9,
        Respvar  =Time,
        Censvar  =VStatus,
        Censval  =0,
        outdsn   =phreg);

*****
**                               PRODUCE SUMMARY TABLE IN MS WORD DOCUMENT                               **
*****

ods rtf file='h:\ibm\phreg.doc';
title1 'Univariate PHREG for endpoint status (Alive, Dead)';
proc report data=phreg box nowd split=' ';
  column Independent Estimate StdErr ProbChiSq HazardRatio
         HRLowerCL HRUpperCL;
  define Independent / display;
  define Estimate    / display;
  define StdErr      / display;
  define ProbChiSq   / display;
  define HazardRatio / display;
  define HRLowerCL   / display;
  define HRUpperCL   / display;

  compute probchisq;
  if probchisq <= 0.05 then
    call define('_c4_', 'style', 'style=[font_weight=bold
    font_style=italic]');
  endcomp;
run;
ods rtf close;

```

OUTPUT

Univariate PHREG for endpoint status (Alive, Dead)

Independent	Parameter Estimate	Standard Error	Pr > Chi-Square	Hazard Ratio	95% Lower Confidence Limit for Hazard Ratio	95% Upper Confidence Limit for Hazard Ratio
LOGBUN	1.74595	0.60460	0.0039	5.731	1.752	18.745
HGB	-0.12494	0.05597	0.0256	0.883	0.791	0.985
PLATELET	-0.69366	0.39665	0.0803	0.500	0.230	1.087
AGE	-0.00216	0.01595	0.8924	0.998	0.967	1.030
LOGWBC	0.56940	0.75489	0.4507	1.767	0.402	7.760
FRAC	0.37140	0.39081	0.3419	1.450	0.674	3.119
LOGPBM	0.33163	0.43415	0.4450	1.393	0.595	3.263
PROTEIN	0.00794	0.02075	0.7020	1.008	0.968	1.050
SCALC	0.10592	0.10041	0.2915	1.112	0.913	1.354

CONCLUSION

Valuable information can be lost in pages of output during the exploratory analysis stage. Biostatistician needs to focus attention to the key information that is primarily of interest before moving on to the next step of analysis. ODS is extremely useful in displaying selected pieces of output. ODS makes this SAS[®] applications a snap. Researchers will appreciate the ability to see a concise table representation of the data in a customized tables and reports. However, caution needs to be exercised especially in the case of dummy variables. Such covariates can be either accommodated appropriately by updating the code to suite the problem at hand or by running individual models for the set of dummy variables of interest.

RECOMMENDED READING

1. Hosmer, D.W., Jr., and Lemeshow, S. (2000), Applied Logistic Regression, New York: John Wiley & Sons, Inc.
2. Krall, J.M., Uthoff, V.A., and Harley, J. B. (1975), "A Step-up Procedure for Selecting Variables Associated with Survival," *Biometrics*, 31, 49 -57.
3. SAS[®] online documentation, SAS[®] Inc, North Carolina.
4. Gupta, S.K., "Using Styles and Templates to Customize SAS ODS Output," SUGI 27 Paper 7-27.
5. Haworth, L., "SAS[®] Reporting 101: Report, Tabulate, ODS, and Microsoft Office," SUGI 28 Paper 71-28.
6. Gupta, S.K., "Quick Results with the Output Delivery System," SUGI 27 Paper 58-27

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Angelina Tan
 Mayo Clinic, Division of Biostatistics
 200 First Street SW
 Rochester MN 55905
 Phone: 507-284 5743
 Fax: 507-266 2477
 Email: tan.angelina@mayo.edu

SAS[®] and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration. Other brand and product names are trademarks of their respective companies.