ODDS RATIOS IN A TABULAR PRESENTATION

Stephen M. Noga, UNC-CH, Chapel Hill, NC
Ding Yi Zhao, UNC-CH, Chapel Hill, NC

Abstract

Odds ratios and their corresponding confidence intervals are easily obtainable statistics using the 'RL' option in PROC LOGISTIC (SAS®/STAT). However, there is no option to output these statistics to a SAS dataset. This presents a challenge to the user when the task at hand is to present those statistics in tabular form, such as output from PROC TABULATE. This paper uses PROC LOGISTIC, PROC TABULATE, and the SAS Macro language to produce the desired tables.

Introduction

So you want to create a table of odds ratios and their 95% confidence intervals, but you only want some of the independent variables listed in that table, and you think you might like to see the results from a number of different models adjacent to each other in this table. On the surface, this does not seem to be an overwhelming problem. After all, that is what PROC TABULATE is for. The $1.98 question is: "How do I get the necessary statistics from PROC LOGISTIC into PROC TABULATE?".

The SAS code to answer that question follows along with a running commentary. This paper is intended for all levels of SAS users. The code was written in version 6.10 of SAS using the OS/2 operating system, but it is not system dependent.

Onward

Let us first look at the output created by using the 'RL' option in the PROC LOGISTIC model statement.

```
PROC logistic data=man.uc24 ;
  model MMSA_OX = PD PH PDH PDD PHH PDDH PHHD
                  PDDHH V1AGE01 SEX D1 D2 D3 D4 / rl;
RUN;
```

This produces the following output:

Conditional Odds Ratios and 95% Confidence Intervals

Variable	Unit	Odds Ratio	Wald Confidence Limits Lower	Upper
PD	1.0000	1.536	1.013	2.329
PH	1.0000	1.422	1.058	1.911
PDH	1.0000	2.481	1.743	3.530
PDD	1.0000	3.831	1.304	11.258
PHH	1.0000	2.280	1.484	3.501
PDDH	1.0000	6.357	2.813	14.362
PHHD	1.0000	4.446	2.820	7.009
PDDHH	1.0000	7.134	2.661	19.123
V1AGE01	1.0000	1.128	1.106	1.152
SEX	1.0000	1.035	0.826	1.296
D1	1.0000	0.947	0.661	1.356
D2	1.0000	3.463	2.498	4.802
D3	1.0000	1.592	1.151	2.202
D4	1.0000	2.705	1.476	4.956

This output has everything you need, but your degree of satisfaction is low because the default output from PROC LOGISTIC is not quite what you need for your presentation.

On the following pages, you will see code for the macro which is used for all calculations and code that calls the macro and produces the tables. The first table simply replicates the PROC LOGISTIC output using PROC TABULATE to pretty it up. The second and third tables expand on this by having multiple models across the table and multiple categories down the table.

The Macros

```
*---------------------------------------------------------------*
| Title:       LOGISTIC                                          |
|                                                                |
| Description: This program uses SAS PROC LOGISTIC to calculate  |
|              odds ratios and 95% confidence limits for each    |
|              explanatory variable.                             |
|                                                                |
| Programmer:  Dingyi Zhao                                       |
| Modified By: Steve Noga                                        |
|              CSCC                                              |
|              Department of Biostatistics                       |
|              University of North Carolina at Chapel Hill       |
|----------------------------------------------------------------|
| You have to provide the following macro variables.            |
|                                                                |
| 1. %let x_num  = K             how many explanatory variables  |
| 2. %let indata = ******        analysis data set.              |
| 3. %let var_y  = Y             response variable(s).           |
| 4. %let var_x  = X1 X2 ... Xk  K explanatory variable(s).      |
| 5. %let drpvar = +++++++       variables to be dropped         |
|                                 from table                     |
|----------------------------------------------------------------|
| The following is the process for calculating odds ratios and  |
| 95% confidence limits.                                         |
|                                                                |
| The first, using PROC LOGISTIC with OUTEST and COVOUT options. |
|                                                                |
| PROC LOGISTIC DATA = &INDATA OUTEST = TEMP COVOUT              |
|      MODEL &VAR_Y = &VAR_X                                     |
|                                                                |
| AFTER RUNNING, ONE GETS VARIANCE AND COVARIANCE MATRIX.        |
|                                                                |
|  _NAME_     INTERCPT     X1    X2     ......   Xk             |
|  ESTIMATE                B1    B2     ......   Bk             |
|    X1                  COV11 COV12   ......  COV1k            |
|    X2                  COV21 COV22   ......  COV2k            |
|    .                      .     .  .   .  .  .                 |
|    Xk                  COVk1 COVk2   ......  COVkk            |
|                                                                |
| LET  B(i) = Bi          i = 1,2, ... , k                       |
|     VAR(i) = COVii      I = 1,2, ... , k                       |
|                                                                |
| THEN ODDS RATIOS FOR EACH EXPLANATORY VARIABLE IS              |
|      OR(i)  = EXP(B(i))             i = 1,2, ... , k           |
|                                                                |
| 95% CONFIDENCE LIMITS                                          |
|     L(i) = EXP(B(i) - 1.96*SQRT(VAR(i)))                       |
|     U(i) = EXP(B(i) + 1.96*SQRT(VAR(i)))                       |
|                                                                |
| WALD STATISTIC FOR EACH EXPLANATORY VARIABLE X(i) IS           |
|  WALD(i) = B(i)*B(i)/VAR(i).  CHI SQUARE WITH 1 DEGREE FREEDOM.|
|  P(i) = 1 - PROCCHI(WALD(i),1) P VALUE FOR EACH X(i).         |
*---------------------------------------------------------------* ;
/******************************
 * TRAN (called by LOGISTIC)  *
 * Gets The Variance And Mean *
 ******************************/
%MACRO tran;
      %do i = 1 %to &x_num;
          %let colnum = %eval(&i+1);
          %let vars= %scan(&xvar,&i);
          if _name_ = "&vars" then do;
             variance = col&colnum;
             mean     = col1;
          end;
      %end;
%MEND tran;
```

```
/************
 * LOGISTIC *
 ************/

   /** change next line for TABLES 2 and 3 **/
 %MACRO Logistic(x_num=,indata=,yvar=,xvar=,drpvar=);
   PROC logistic data=&indata outest=temp covout;
    model &yvar = &xvar;
   RUN;
   /**********************************************
    * Take output dataset from proc logistic and: *
    * Rename and possibly drop variables          *
    * Delete certain observations                 *
    **********************************************/
   DATA temp;
     set temp(keep=_name_ &xvar
              rename=(_name_=variable) drop=&drpvar);

   /*************************************************
    * add any variables that you want deleted, below *
    *************************************************/
   if "&drpvar" gt ' ' then do;
     if variable = 'INTERCPT' or
        index("&drpvar",trim(variable)) > 0 then delete;
   end;
   else do;
     if variable = 'INTERCPT' then delete ;
   end;
   RUN;

   PROC transpose data=temp out=model;
   RUN;

   /************************************************
    * Take output dataset from proc transpose and: *
    * Rename variable                              *
    * Call macro TRAN to compute mean and variance *
    * Compute ORs, Upper/Lower 95% CIs             *
    * Compute Parameter Estimate and P Value       *
    ************************************************/
   DATA good(rename=(_name_=variable));
   /** add length statement here for TABLES 2 and 3 **/
     set model;
   %tran;
   L95_OR  = exp(mean-1.96*sqrt(variance));
   U95_OR  = exp(mean+1.96*sqrt(variance));
   OR      = exp(mean);
   B_VALUE = mean;
   WALD    = (mean*mean)/variance;
   p       = 1 - probchi(wald,1);
   if p < 0.0005 then P_WALD = 0.001;
    else P_WALD = round(p,0.001);
   /** add assignment statements here for
       TABLES 2 and 3 **/
   output;
   RUN;

   /** add proc append here for TABLES 2 and 3 **/
 %MEND Logistic;
```

Code For TABLE 1

```
PROC format;
 value $pfmt
  'PD'    = '        1 x DIAB'
  'PH'    = '       1 x HTN'
  'PDH'   = '      1 x DIAB and 1 x HTN'
  'PDD'   = '     2 x DIAB'
  'PHH'   = '    2 x HTN'
  'PDDH'  = '  2 x DIAB and 1 x HTN'
  'PHHD'  = ' 2 x HTN and 1 x DIAB'
  'PDDHH' = '2 x DIAB and 2 x HTN'   ;
RUN;

DATA job;
  set man.uc24
   (keep=MMSA_OX PD PH PDH PDD PHH PDDH PHHD PDDHH
         V1AGE01 SEX D1 D2 D3 D4 WSTHPRO1 BMI01) ;
RUN;
/*****************************************
 *  Call macro LOGISTIC which computes    *
 *  the odds ratios and CIs               *
 *  indata = Input Dataset                *
 *  x_num  = # of Independent Variables   *
 *           you want to See in Output    *
 *  y_var  = Dependent Variable           *
 *  x_var  = Independent Variables        *
 *  drpvar = Variables that get Dropped   *
 *****************************************/
%Logistic(indata=job,
          x_num=8,
          yvar=MMSA_OX,
          xvar=PD PH PDH PDD PHH PDDH PHHD PDDHH V1AGE01 SEX D1 D2 D3 D4,
          drpvar=V1AGE01 SEX D1 D2 D3 D4);

/*******************
 * Produce Table 1 *
 *******************/
 PROC tabulate data=good noseps order=formatted;
   CLASS variable;
   VAR or l95_or u95_or ;
   TABLE variable = ' ',
         (or l95_or u95_or)*sum=' '*f=10.3 /
         BOX='Number and Type of Disorders*' rts=25;
   LABEL or      = 'Odds Ratio'
         l95_or  = '95% CI Lower Bound'
         u95_or  = '95% CI Upper Bound';
   FORMAT variable $pfmt. ;
   TITLE2 'TABLE 1';
   FOOTNOTE '* Diabetes (DIAB) and Hypertension (HTN)';
 RUN;
```

Code For TABLE 2

Now that we know we can replicate the results from PROC LOGISTIC, we want to increase the information shown in the table. To do that, we need to add a few lines of code to %LOGISTIC and modify a few other lines.

The macro variable **MD** (model number) needs to be added to the list of variables.

```
%MACRO Logistic(md=,indata=,x_num=,yvar=,xvar=,drpvar=);
```

Then we need to specify the length of **MODNUM** after the "DATA good...." line.

```
length modnum $6 ;
```

Assign the value of **&MD** to **MODNUM**. This is inserted right before the "output" statement in the "DATA good" section.

```
modnum = "&md" ;
```

Need to append all models into the dataset "all", so you have to insert these lines right after the "DATA good" section.

```
        /*********************
         * Append All Models *
         *********************/
        PROC APPEND base=all DATA=test  ;
        run; quit;
```

Those are all the changes and modifications to %LOGISTIC that you need to create TABLE 2. Now let's look at how the code that calls %LOGISTIC needs to be tweaked to produce the desired table.

Since you are appending datasets, it is a good idea to make sure that the dataset "all" does not already exist. Run the following procedure to take care of that.

```
        PROC datasets lib=work;
         delete all;
        RUN; QUIT;
```

Then you need to add the variable **md** to your macro call.

```
/*******************************
 *  Additional Variable Needed *
 *  md = Model Number          *
 *******************************/
%Logistic(md=1**,indata=job,x_num=8,yvar=mmsa_0x,
          xvar=PD PH PDH PDD PHH PDDH PHHD PDDHH V1AGE01 SEX D1 D2 D3 D4,
          drpvar=V1AGE01 SEX D1 D2 D3 D4);
%Logistic(md=2***,indata=job,x_num=8,yvar=mmsa_0x,
          xvar=PD PH PDH PDD PHH PDDH PHHD PDDHH V1AGE01 SEX D1 D2 D3 D4 BMI01,
          drpvar=V1AGE01 SEX D1 D2 D3 D4 BMI01);
%Logistic(md=3****,indata=job,x_num=8,yvar=mmsa_0x,
          xvar=PD PH PDH PDD PHH PDDH PHHD PDDHH V1AGE01 SEX D1 D2 D3 D4 WSTHPRO1,
          drpvar=V1AGE01 SEX D1 D2 D3 D4 WSTHPRO1);
%Logistic(md=4*****,indata=job,x_num=8,yvar=mmsa_0x,
          xvar=PD PH PDH PDD PHH PDDH PHHD PDDHH V1AGE01 SEX D1 D2 D3 D4 WSTHPRO1 BMI01,
          drpvar=V1AGE01 SEX D1 D2 D3 D4 WSTHPRO1 BMI01);
```

Finally, update your PROC TABULATE code (**highlighted**).

```
*-----------*
¦  TABLE 2  ¦
*-----------*;
PROC TABULATE DATA=all noseps ORDER=FORMATTED;
  CLASS variable modnum;
  VAR or l95_or u95_or ;
  TABLE variable = ' ',
       modnum='Model Number'*( (or l95_or u95_or)*sum=' '*f=10.3
       ) / rts=25 BOX='Number and Type of Disorders*';
  LABEL or     = 'Odds  Ratio'
       l95_or = '95% CI Lower Bound'
       u95_or = '95% CI Upper Bound' ;
  FORMAT variable $pfmt. ;
  TITLE2 '                                                                   TABLE 2';
  FOOTNOTE1 '*     Diabetes (DIAB) and Hypertension (HTN)';
  FOOTNOTE2 '**    Adjusted for Age, Gender, Ethnicity/Center';
  FOOTNOTE3 '***   Adjusted for Age, Gender, Ethnicity/Center, BMI';
  FOOTNOTE4 '****  Adjusted for Age, Gender, Ethnicity/Center, WHR';
  FOOTNOTE5 '***** Adjusted for Age, Gender, Ethnicity/Center, BMI, WHR';
RUN;
```

And that is all that is needed to create TABLE 2.

Code For TABLE 3

Finally, let's look at what is needed to create a table with two variables going down the left side and two models across. The only changes to %LOGISTIC are adding

```
        mm $8
```

to the length statement, and inserting the following line

```
        mm = "&yvar" ;
```

right before the "output" statement, with both of these going into the "DATA good" section.
The only changes/additions needed in the code that calls %LOGISTIC are adding necessary formats,

```
         value $mmfmt
          'MMSA_AOD' = ' MM vs All Other Diabetics'
          'MMSA_AOH' = 'MM vs All Other Hypertensives';
         value $phfmt
          'P1' = '1'
          'P2' = '2'
          'P3' = '3 - 4';
```

changing the variables in the macro call statement to reflect what you want in the table,

```
        %Logistic(md=1*,indata=job,x_num=3,yvar=MMSA_AOD,
                  xvar=P1 P2 P3 V1AGE01 SEX D1 D2 D3 D4,
                  drpvar=V1AGE01 SEX D1 D2 D3 D4);
        %Logistic(md=1*,indata=job,x_num=3,yvar=MMSA_AOH,
                  xvar=P1 P2 P3 V1AGE01 SEX D1 D2 D3 D4,
                  drpvar=V1AGE01 SEX D1 D2 D3 D4);
        %Logistic(md=2**,indata=job,x_num=3,yvar=MMSA_AOD,
                  xvar=P1 P2 P3 V1AGE01 SEX D1 D2 D3 D4 BMI01 WSTHPR01,
                  drpvar=V1AGE01 SEX D1 D2 D3 D4 WSTHPR01 BMI01);
        %Logistic(md=2**,indata=job,x_num=3,yvar=MMSA_AOH,
                  xvar=P1 P2 P3 V1AGE01 SEX D1 D2 D3 D4 BMI01 WSTHPR01,
                  drpvar=V1AGE01 SEX D1 D2 D3 D4 WSTHPR01 BMI01);
```

and finally modify the PROC TABULATE.

```
          CLASS mm variable modnum;
          TABLE mm=''*
                variable = '',
                modnum='Model Number'*( (or l95_or u95_or)*sum=' '*f=10.3) /
                rts=46 box="Case-Control Contrast  Number Of          By Type Of Component    Histories";
          FORMAT mm $mmfmt. variable $phfmt. ;
         TITLE5 '                                   TABLE 3';
         FOOTNOTE1 '*  Adjusted for Age, Gender, Ethnicity/Center';
         FOOTNOTE2 '** Adjusted for Age, Gender, Ethnicity/Center, BMI, WHR';
```

This creates TABLE 3.

Conclusion

So now you know one way to get odds ratios and confidence intervals into PROC TABULATE output. There are other ways to do the same thing, this approach is just one that is easily modified to create the different table looks. In addition, macro variables can be added to take advantage of the different features of PROC LOGISTIC, but that is for another day.

Author Contact

Stephen Noga   (uccsmn.cscc@mhs.unc.edu)
Dingyi Zhao   (uccdyz.cscc@mhs.unc.edu)
University Of North Carolina - CH
Collaborative Studies Coordinating Center
137 E. Franklin Street
Suite #203
Chapel Hill, NC 27514
(919) 962-6971

TABLE 1

Number and Type of Disorders*	Odds Ratio	95% CI Lower Bound	95% CI Upper Bound
1 x DIAB	1.536	1.013	2.329
1 x HTN	1.422	1.058	1.911
1 x DIAB and 1 x HTN	2.481	1.743	3.530
2 x DIAB	3.831	1.304	11.258
2 x HTN	2.280	1.484	3.501
2 x DIAB and 1 x HTN	6.357	2.813	14.363
2 x HTN and 1 x DIAB	4.446	2.820	7.010
2 x DIAB and 2 x HTN	7.134	2.661	19.123

* Diabetes (DIAB) and Hypertension (HTN)

TABLE 2

Number and Type of Disorders*	Model Number											
	1**			2***			3****			4*****		
	Odds Ratio	95% CI Lower Bound	95% CI Upper Bound	Odds Ratio	95% CI Lower Bound	95% CI Upper Bound	Odds Ratio	95% CI Lower Bound	95% CI Upper Bound	Odds Ratio	95% CI Lower Bound	95% CI Upper Bound
1 x DIAB	1.536	1.013	2.329	1.426	0.924	2.200	1.328	0.855	2.064	1.308	0.836	2.044
1 x HTN	1.422	1.058	1.911	1.216	0.888	1.664	1.337	0.974	1.835	1.215	0.878	1.680
1 x DIAB and 1 x HTN	2.481	1.743	3.530	1.981	1.356	2.895	2.222	1.516	3.257	1.964	1.328	2.905
2 x DIAB	3.831	1.304	11.258	4.623	1.541	13.866	4.155	1.317	13.111	4.656	1.476	14.682
2 x HTN	2.280	1.484	3.501	2.031	1.279	3.227	2.071	1.293	3.319	1.854	1.136	3.025
2 x DIAB and 1 x HTN	6.357	2.813	14.363	4.321	1.704	10.956	5.690	2.375	13.629	4.380	1.676	11.442
2 x HTN and 1 x DIAB	4.446	2.820	7.010	3.925	2.420	6.366	3.679	2.228	6.075	3.406	2.043	5.678
2 x DIAB and 2 x HTN	7.134	2.661	19.123	6.501	2.246	18.814	8.802	3.204	24.181	8.260	2.993	22.797

* Diabetes (DIAB) and Hypertension (HTN)
** Adjusted for Age, Gender, Ethnicity/Center
*** Adjusted for Age, Gender, Ethnicity/Center, BMI
**** Adjusted for Age, Gender, Ethnicity/Center, WHR
***** Adjusted for Age, Gender, Ethnicity/Center, BMI, WHR

TABLE 3

Case-Control Contrast By Type Of Component	Number Of Histories	Model Number					
		1*			2**		
		Odds Ratio	95% CI Lower Bound	95% CI Upper Bound	Odds Ratio	95% CI Lower Bound	95% CI Upper Bound
MM vs All Other Diabetics	1	0.909	0.654	1.265	0.891	0.638	1.245
	2	1.097	0.764	1.577	1.121	0.776	1.621
	3 - 4	1.725	1.061	2.807	1.689	1.030	2.770
MM vs All Other Hypertensives	1	0.980	0.744	1.290	0.987	0.745	1.308
	2	1.380	1.020	1.867	1.410	1.034	1.922
	3 - 4	2.775	1.876	4.105	2.549	1.698	3.826

* Adjusted for Age, Gender, Ethnicity/Center
** Adjusted for Age, Gender, Ethnicity/Center, BMI, WHR