

Applications in SAS/IML™: Getting Odds Ratios from PROC LOGISTIC

D. Kim Chantala, WESTAT, Inc., Durham, NC

Abstract

One limitation of PROC LOGISTIC is the inability to compute odds ratios for combinations of the parameters estimated by the MODEL statement. Certain versions of the SAS™ System will give you the odds ratio for each parameter estimated, but not for the sum or difference of two of the parameters or a multiple of one of the parameter estimates. SAS/IML was used to write a simple routine that solves this problem. Using the covariance matrix output data set from PROC LOGISTIC, this program estimates a user specified linear function of the parameters. Printed output includes:

- * the estimate of the linear combination of the parameters
- * the estimated standard error of the linear combination of the parameters
- * the associated chi-square test statistic and p-value
- * odds ratio and 95% confidence interval.

The code for the program, instructions for its use and an example are presented.

Introduction

Because of its ease of interpretation, the odds ratio is often the desired result from PROC LOGISTIC. Usually the point estimate of the odds ratio as well as estimates of a confidence interval or p-values are needed for presentation in tables from the logistic regression. If the needed odds ratio is simply one of the exponentiated regression coefficients it is quite easy to compute the confidence interval by hand using the standard error printed automatically by PROC LOGISTIC. However, the desired odds ratio may turn out to be a linear combination of the parameter estimates (the regression coefficients). This can occur if interactions are present in your final model. You will have to determine the appropriate formula for the variance of the linear combination describing the desired odds ratio and then obtain the appropriate values for the covariance terms to plug into the derived formula for the standard error. Although PROC LOGISTIC will optionally print the variance-covariance matrix to facilitate hand calculations, it is much easier to use SAS/IML

(Interactive Matrix Language) to compute the matrix algebra formulas for the variance, confidence intervals, odds ratio and p-values. The advantage of using matrix representation is that the matrix form of the equation used to compute the variance is always the same, where the algebraic form must always be derived any time the linear combination changes.

SAS/IML treats matrices as objects, freeing the user from keeping track of the attributes associated with each matrix. Capabilities of this powerful language include creating and manipulating data sets, a variety of matrix operators and functions, a wide range of graphics plus many of the same features for character and arithmetic expressions and structured programming that are available in the BASE SAS Software. The program included in this paper illustrates several of these features.

Formulas in Matrix Representation

Only two formulas from matrix algebra are needed: one for the log of the odds ratio (log odds) expressed as a linear combination of the parameter estimates and one for the variance of the linear combination.

In general, the procedure to calculate the log odds uses the predicted model to calculate the difference between the dependent variable evaluated at two different levels of the independent variable(s). This difference expresses the desired log odds as a linear combination of the parameter estimates and can be expressed as the product of two vectors:

$$\ln(\text{odds}) = L'b$$

where

- b** is a vector containing the parameter estimates and will be obtained from PROC LOGISTIC
- L'** is the transpose of a vector that specifies the linear combination of the parameter estimates and will be derived by the user.

The variance of the log odds is calculated by:

$$\text{var}\{\ln(\text{odds})\} = L'VL$$

where **V** is the variance-covariance matrix for the parameter estimates and is obtained from PROC LOGISTIC.

Implementing the Formulas in IML

This section describes several features of SAS/IML used in the program ODDSRATIO.SAS listed in Appendix A. The following section titles correspond to identical comments in ODDSRATIO.SAS and will aid you in locating the SAS/IML statements being discussed.

/* ELIMINATE UNNECESSARY VARIABLE */

The variance-covariance matrix and parameter estimates were obtained by using the COVOUT OUTEST=BCOV options with the PROC LOGISTIC statement when you fit your final model. You will do this prior to using ODDSRATIO.SAS. Instructions are given in the section titled "How to use ODDSRATIO.SAS". The output data set, BCOV, contains the names, estimates, and covariances of the intercept and the explanatory variables specified in the MODEL statement. An unnecessary variable, **_LNLIKE_**, is dropped in a data step creating NEW_BCOV at the beginning of the source code in ODDSRATIO.SAS.

/* CREATE VECTORS b & L, MATRIX V */

By using SAS/IML statements we can extract portions of this data set and save them as the matrices needed to do the calculations. The USE statement opens the data set NEW_BCOV and makes it the default data set for any subsequent statements. The READ statement allows you to extract portions of the data set to be stored as matrices. Special options allow you to select all numeric variables (**ALL VAR _NUM_**) that satisfy special conditions (**WHERE _TYPE_="PARAMS"**). The variable **_TYPE_** stored in the data set NEW_BCOV was supplied by PROC LOGISTIC and has either the value **PARAMS** to denote the observation containing values for the parameter estimates or the value **COV** to denote the observations containing values for the covariance matrix. By using the READ statement the parameter vector, **b**, and the covariance matrix, **V**, are extracted from the data set NEW_BCOV while the linear combination vector, **L**, is extracted from the data set LVEC. Details on creating LVEC are given in the section 'How to Use ODDSRATIO.SAS'.

/* MATRIX CALCULATIONS & ATTRIBUTES */

Since IML keeps track of the dimensions of each matrix, the mathematical operators apply to the entire matrix and simplify the programming task. Transposing a matrix is achieved by following the name of the matrix by the symbol **`** while the **\*** operator instructs IML to multiply two matrices (**LNOR = L` * b**). Similarly the **+** operator will add two matrices. Two matrices can be concatenated horizontally by the operator **||**. Hence, the statement **LOW95 || HI95** uses the upper (HI95) and lower (LOW95) bounds of the 95% confidence interval to create a matrix with 1 row and 2 columns. The MATTRIB statement allows the user to define a label (the printing attributes of a matrix) for each row or column in a matrix. Each time the matrix is printed, these labels are printed by the corresponding row or column.

How to use ODDSRATIO.SAS

This section uses a hypothetical example to illustrate the use of ODDSRATIO.SAS. The code to access the statements in ODDSRATIO.SAS is shown at the

end of this section. The output appears in Appendix B.

Determine Vector L

Suppose the best model to predict the presence or absence of a disease uses the following covariates:

AGE (a continuous variable)
SEX (0=female, 1=male),
drug A (0=not administered/1=administered)
the interaction of SEX and drug A, SEX_A.

To compare the odds of getting the disease and being male while taking drug A to being female and taking drug A you would compute the difference in the model evaluated under each of these two conditions:

$$\begin{aligned} \ln(\text{odds}) &= \\ & [\text{INTERCEPT} + \text{AGE} \cdot \text{years}_i + \text{SEX} \cdot 1 + A \cdot 1 + \text{SEX}_A \cdot 1] \\ & - [\text{INTERCEPT} + \text{AGE} \cdot \text{years}_i + \text{SEX} \cdot 0 + A \cdot 1 + \text{SEX}_A \cdot 0] \\ & = \text{INTERCEPT} \cdot 0 + \text{AGE} \cdot 0 + \text{SEX} \cdot 1 + A \cdot 0 + \text{SEX}_A \cdot 1 \\ & = L' b, \text{ where } L' = (0, 0, 1, 0, 1). \end{aligned}$$

Code to Use ODDSRATIO.SAS

Listed below is the code to compute the odds ratio for our example. There are two restrictions on using this program: the input data sets must be named BCOV (output data set from PROC LOGISTIC COVOUT OUTEST=BCOV) and LVEC (data set that specifies the desired linear combination of the parameter estimates). It is important that the first element of L' corresponds to the INTERCEPT and the following elements must be in the same order as the covariates

are entered on the model statement. The %include statement allows you to store the code for ODDSRATIO.SAS in an external file, accessing its contents when needed.

```
options source2;
proc logistic data=mydata covout outest=bcov;
model FCLASS = AGE SEX A SEX_A ;
run;

/**** The following code creates a data set called
LVEC from the vector X. You can substitute a data
step to create the data set LVEC ****/

proc iml;
L={0,0,1,0,1};
create lvec var {L} ;
append var {L};
run;
title3 'Output from ODDSRATIO.SAS';
%include 'oddsratio.sas';
```

The output appears in Appendix B gives the estimated value of the odds ratio, p-value, and 95% confidence interval.

Conclusions

The use of ODDSRATIO.SAS simplified analysis and shortened the turn around time for getting results to our investigators. Possible errors caused by doing calculations by hand have been eliminated. The code would be more flexible if modified to allow a user to select any name for the data sets BCOV and LVEC.

Appendix A: Source Code to compute Odds Ratio

```
*****
PROGRAM: ODDSRATIO.SAS

FUNCTION: Calculates Odds Ratios from a vector specifying the linear
          combination of parameters and a covariance matrix.

LANGUAGE: SAS/IML

SYSTEM: VAX/VMS

TO USE:
1. Create a data set called LVEC that specifies the desired linear
   combination of the parameter estimates.
2. Use the COVOUT OUTEST=BCOV options on the PROC LOGISTIC
```

statement.

3. Use %INCLUDE 'ODDSRATIO.SAS' to access the statements stored in this file named ODDSRATIO.SAS.

RELATED DOCUMENTS:

SAS Institute Inc., "SAS/IML Software: Usage and Reference, Version 6", First Edition, Cary, NC: SAS Institute Inc., 1989, 501 pp.

Hosmer, David W. and Lemeshow, Stanley, "Applied Logistic Regression", New York, John Wiley & Sons, 1989, 307 pp.

KNOWN RESTRICTIONS:

The input data sets must be named BCOV (output data set from PROC LOGISTIC COVOUT OUTEST=BCOV) and LVEC (vector that specifies the desired linear combination of the parameter estimates).

*****/

/* ELIMINATE UNNECESSARY VARIABLE */

```
data new_bcov;
set bcov;
drop _lnlike_;
```

```
;
proc iml;
```

/* CREATE VECTORS b & L, MATRIX V */

```
use new_bcov;
read all var _num_ into param where (_type_="PARMS");
read all var _num_ into V where (_type_="COV");
```

```
use lvec;
read all into l;
```

```
b=param';
```

* (Uncomment the following if you want to see the input values);

```
*print (b') " = Parameter vector";
*print (L') " = L - linear combination of parameters";
*print V;
```

/* MATRIX CALCULATIONS & ATTRIBUTES */

```
lnor = L'*b;          mattrib lnor rowname = 'LOG ODDS =';
or = exp(lnor);       mattrib or rowname = 'ODDS RATIO =';
var = L'*V*L;        mattrib var rowname = 'Variance LOG ODDS =';
se = sqrt(var);      mattrib se rowname = 'Standard Error LOG ODDS =';
chi = (lnor/se)**2;   mattrib chi rowname = 'Chi-Square';
p_value = 1-probchi(chi,1); mattrib p_value rowname = 'Pr > Chi-Square';
low95 = exp(lnor-1.96*se); * lower bound of Confidence interval;
hi95 = exp(lnor+1.96*se); * upper bound of Confidence interval;
CI95 = low95||hi95;   mattrib CI95 rowname = 'ODDS RATIO 95% CI';
```

/* PRINT RESULTS */

```
print lnor;
print var;
print se;
print chi;
print p_value;
print or;
```

```
print CI95;
run;
```

Appendix B: Output

The output from PROC LOGISTIC has been omitted.

```
Output from ODDSRATIO.SAS

LNOR
LOG ODDS = -0.133528

VAR
Variance LOG ODDS = 0.0033518

SE
Standard Error LOG ODDS = 0.0578946

CHI
Chi-Square 5.3194985

P_VALUE
Pr > Chi-Square 0.0210881

OR
ODDS RATIO = 0.8750027

CI95
ODDS RATIO 95% CI 0.7811394 0.9801449
```

References

SAS Institute Inc., (1989), *SAS/IML Software: Usage and Reference, Version 6, First Edition*, Cary, NC: SAS Institute Inc., 1989, 501 pp.

Hosmer, David W. and Lemeshow, Stanley, (1989), *Applied Logistic Regression*, New York: John Wiley & Sons, 307 pp.

SAS and SAS/IML are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries.

VAX and VMS are registered trademarks or trademarks of Digital Equipment Corporation in the USA and other countries.