Logistic Regression Adjustment of Proportion and its Macro Procedure

Dingyi Zhao
UNC-CH, Chapel Hill, NC

Abstract

Logistic regression model can be used to calculate the conditional probability that outcome is present as denoted by \( P(Y=1|X) \). To adjust some confounding factors, one can consider using the adjusted proportions. This paper introduces an adjustment method based on logistic regression model and presents its corresponding macro procedure. The macro presented here enables to calculate multivariate adjusted proportions and its corresponding confidence intervals and outputs the results in a tabular form.

Method

Let the conditional probability that outcome is present as denoted by \( P(Y=1|X) \). Then the logit of the logistic regression model is given by the equation:

\[
\text{Logit}(P(Y=1|X)) = b_0 + b_1 x_1 + \ldots + b_i x_i + \ldots + b_p x_p;
\]

Consider relationship between \( Y \) with \( x_i \) after controlling the rest of \( X \) variables, where \( x_i = 0 \) and 1 value.

Assume \( \text{betas} = (\beta_0, \beta_1, \ldots, \beta_p) \) estimators, respectively.

\[
\text{MX} = (1 \ x_1 \ \ldots \ 0 \ \ldots \ \text{mxp,} \ \ 1 \ x_1 \ \ldots \ 1 \ \ldots \ \text{mxp)}
\]

where \( \text{mxj} \) is mean of \( x_j \) (\( j=1, \ldots, p \) and not equal \( i \)).

Then, \( P \) at mean of \( X \) is

\[
P(Y=1|\text{MX}) = \exp(\text{betas} \times \text{MX'}) / (1 + \exp(\text{betas} \times \text{MX'}));
\]

and its 95% CI is

\[
95\%\ CI = \exp(ci) / (1 + \exp(ci));
\]

where \( ci = \text{betas} \times \text{MX'} + 1.96 * \text{sqrt}(\text{VERDIAG} (\text{MX} \times \text{COV} \times \text{MX'}));
\]

\( \text{COV} \) is variance–covariance matrix. Due to nonlinear nature of the logistic model, \( P \) at mean of \( X \) is not equal to mean of \( P \) over the sample, i.e. \( f(E(x)) \neq E(f(x)) \). So a corrected coefficient, \( K \), have to be used:

\[
K = \frac{\text{Actual Proportions overall}}{\text{predicted proportion overall}} = \frac{P(Y=1)}{((p_0 \ p_1) \times (P(Y=1|\text{MX}'))'};
\]

where \( p_i \) are the proportions of \( x_i = 0 \) and 1, respectively.

The corrected proportion and its 95% CI are

\[
P_c = K \times P(Y=1|\text{MX});
\]

\[
95\% \text{CI}_c = K \times (95\% \text{CI});
\]

Macro Procedure

The macro named as ADJ_PROP was written in version 6.10 of SAS using the OS/2 operating system, in which includes PROC LOGISTIC and IML. The macro SAS codes are as below:

<table>
<thead>
<tr>
<th>Macro Name: ADJ_PROP</th>
</tr>
</thead>
<tbody>
<tr>
<td>VERSION : 1</td>
</tr>
<tr>
<td>FUNCTION : MACRO ADJ PROP CAN BE USED TO CALCULATE ADJUSTED MEANS AND PROPORTIONS.</td>
</tr>
</tbody>
</table>

Macro Parameters:
1. INFILE = INPUT DATASET
2. MODEL = REG OR LOGISTIC
3. BYVAR = BY VARIABLE
4. YVARS = Y VARIABLES (VARIABLES SEPARATED BY BLANK)
5. IX VAR = INTERESTED X VARIABLES (SEPARATED BY BLANK)
6. CONTVARS = CONTROL VARIABLES (SEPARATED BY BLANK)
7. GIVENVAR = CONTROL VARIABLE WITH A GIVEN REFERENCE VALUE
8. GIVENVAL = GIVEN REFERENCE VALUE IF CONTVAR2 CHOSEN
9. PRINT = YES, PRINT OUT
INTERMEDIATE RESULTS.

10. DES = DES, IF LOGISTIC AND Y=1 FOR OUTCOME PRESENT

11. OUT = OUTPUT DATASET

MACRO FORMAT:
%ADJ_PROP(INFILE,MODEL,BYVAR,YVARS,IX_VAR,CONTVARS,GIVENVAR,GIVENVAR,PRINT,DES,OUT)

EXAMPLE USAGES:
%ADJ_PROP(ONE,REG,GENDER,HDL,STIA,FB,AGE,55,YES,DES,OUT)
%ADJ_PROP(TWO,LOGISTIC,GENDER,CHD,STIA,FB,AGE,55,YES,DES,OUT)

AUTHOR: DINGYI ZHAO
DEPARTMENT OF BIOSTATISTICS
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL.
DATE: 09/17/96

%MACRO ADJ_PROP(INFILE,MODEL,BYVAR,YVARS,IX_VAR,CONTVARS,GIVENVAR,GIVENVAR,PRINT,DES,OUT);
%GLOBAL VAR_NUM;

*--------------------------*
| COUNT NUMBER OF VARIABLES |
*--------------------------*

%MACRO COUNT(COUNTN,COUNTVAR);
%GLOBAL &COUNTN.NUM;
%IF %LENGTH(&COUNTVAR)>0 %THEN %DO;
%LET COUNT1=1;
%DO %UNTIL(%SCAN(&COUNTVAR,&COUNT1)=) ;
%LET COUNT1=%EVAL(&COUNT1+1);
%END;
%LET &COUNTN.NUM = %EVAL(&COUNT1-1);
%PUT &COUNTN.NUM;
%END;
%MEND;

%COUNT(BY,&BYVAR);
%COUNT(CONT,&CONTVARS);
%COUNT(IX,&IX_VAR);
%COUNT(Y,&YVARS);

*----------------------------*
| FIND NUMBER OF ALL EXPLANATORY |
| VARIABLES.                    |
*----------------------------*

%IF &GIVENVAR NE %THEN %DO;
%LET VAR_NUM = %EVAL(&COUNT_NUM +&IX_NUM+2);
%END;
%ELSE %DO;

%LET VAR_NUM = %EVAL(&CONT_NUM +&IX_NUM+1);

%END;
%LET TAB_NUM=%EVAL((&IX_NUM+1)*12);
%PUT VAR_NUM = &VAR_NUM;
%PUT TAB_NUM = &TAB_NUM;

*--------------------------*
| TRANSLATE 'X1 X2' TO |
| X1="*X2=".          |
*--------------------------*

%MACRO TABFACT(NUMT,TABF,TVAR);
%IF %LENGTH(&TVAR)>0 %THEN %DO;
%GLOBAL &TABF.TAB;
%LET TAB = ;
%DO DI = 1 %TO &NUMT;
%IF &DI NE &NUMT %THEN %DO;
%LET IXADD = %SCAN(&TVAR,&DI)="*";
%END;
%ELSE %DO;
%LET IXADD =
%END;
%LET &TABF.TAB = &TAB;
%END;
%MEND;

%TABFACT(&IX_NUM,IX,&IX_VAR);
%TABFACT(&BY_NUM,BY,&BYVAR);

*----------------------------*
| TRANSLATE A STRING 'X1 X2 |
| X3' TO X1*X2*X3          |
*----------------------------*

%MACRO TRANS(XX,TT);
%IF %LENGTH(&XX)>0 %THEN %DO;
%GLOBAL &TT;
CALL SYMPUT('"&TT", TRANSLATE("&XX","*",""));
%END;
%MEND;

DATA _NULL_;
%TRANS(&IX_VAR,IXVAR);
%TRANS(&BYVAR,BYSORT);
RUN;

*--------------------------*
| DO LOOP Y VARIABLES    |
*--------------------------*

%DO YI = 1 %TO &Y_NUM;
%LET YVAR=%SCAN(&YVARS,YI);

*--------------------------*
| SELECT REG OR LOGISTIC |
| MODEL.                 |
*--------------------------*
%IF %UPCASE(&PRINT) NE YES %THEN
  %DO;
  %LET OUTPRT = NOPRINT;
  %END;
%ELSE %DO;
  %LET OUTPRT = ;
%END;
PROC &MODEL DATA=&INFILE &DES
  OUTEST=STATS COVOUT &OUTPRT;
  %IF %LENGTH(&BYVAR)>0 %THEN
  %DO;BY &BYVAR;
  %END;
MODEL &YVAR = &GIVENVAR &CONTVARS &IX_VAR;
RUN;
%IF %UPCASE(&PRINT) = YES %THEN %DO;
  PROC PRINT DATA = STATS;
  RUN;
%END;
*---------------------------- *
| CALCULATE PROPORTIONS OR |
| MEANS OF CONTROL VARIABLES |
| FROM ALL SAMPLES.          |
| *__________________________* |
%IF %LENGTH(&CONTVARS)>0 %THEN
  %DO;
  PROC MEANS DATA = &INFILE NOPRINT;
  VAR &CONTVARS;
  OUTPUT OUT= PROP
    MEAN=PROP1- PROP&CONT_NUM;
  RUN;
  %LET KEEPPROP=PROP1-
    PROP&CONTVARS;
  %END; %ELSE %DO;
  %LET KEEPPROP = ;
%END;
*---------------------------- *
| CALCULATE SUM OF Y VARIABLE |
| FROM ALL SAMPLES.            |
| *__________________________* |
%IF %UPCASE(&MODEL) = LOGISTIC %THEN
  %DO; PROC MEANS DATA = &INFILE NOPRINT;
    VAR &YVAR;
    OUTPUT OUT= YSUM SUM=YNUMBER;
  RUN;
%END;
*---------------------------- *
| FORM A COEFFICIENT MATRIX FOR |
| CONTROL VARIABLES              |
| *__________________________________________________________* |
DATA ADJ_COEF(KEEP=INT &KEEPPROP &GIVENVAR);
  IF _N_ = 1 THEN DO;
  %IF %LENGTH(&CONTVARS)>0 %THEN
    %DO; SET PROP(KEEP=KEEPPROP);
    END;
    %IF &GIVENVAL> .Z %THEN DO;
      &GIVENVAR= &GIVENVAL;
    END;
    INT = 1;
    RUN;
%END;
*---------------------------- *
| DETERMINE HOW MANY LEVELS IN |
| BY VARIABLE.                 |
| *__________________________* |
%IF %LENGTH(&BYVAR)>0 %THEN %DO;
  PROC FREQ DATA = &INFILE NOPRINT;
    TABLE &BYSORT/OUT=BYL;
  RUN;
  DATA NULL ;
  SET BYL(KEEP=&BYVAR) END=LAST;
  %DO BI = 1 %TO &BY_NUM;
    %LET BYV=%SCAN(&BYVAR, &BI) ;
    CALL SYMPUT("BY&BI", LEFT(_N_), TRIM(&BYV));
  %END;
  IF LAST THEN CALL SYMPUT('BYLEVEL', _N_);
  RUN;
%END;
%ELSE %DO;
  %LET BYLEVEL = 1;
%END;
%MACRO CONTS;
%IF %LENGTH(&CONTVARS)>0 %THEN
  %DO;
    %DO J = 1 %TO &CONTVARS;
      PROP&J
    %END;
  %END;
%MEND ;
%MACRO IXS;
%DO KI= 1 %TO &IX_NUM;
  IX&KI
%END;
%MEND ;
*---------------------------- *
| MATRIX OPERATION BY USING |
| SAS IML.                    |
| *----------------------------* |
PROC IML;
%IF %UPCASE(&PRINT) = YES %THEN %DO;
  %DO; RESET PRINT;
  %END;
%END;
ONE = I(&IX_NUM);
ZERO = SHAPE(0,1,&IX_NUM);
INIT = ZERO//ONE;
PROP={INT &GIVENVAR %CONTS};
IX={&IXS};
XVAR={INTERCEP &GIVENVAR &CONTVARS &IX_VAR};
USE ADJ_COEF;
READ ALL VAR PROP INTO PX;
CLOSE;
X = PX||INIT;

*-----------------------------*
| BETAS & COVS FROM MODEL |
*-----------------------------*

USE STATS;
READ ALL VAR XVAR INTO BETA
WHERE(_TYPE_="PARMS") ;
READ ALL VAR XVAR INTO COVS
WHERE(_TYPE_="COV") ;
CLOSE;

*------------------------------*
| CALCULATE PREDICTED |
| PROPORTIONS AND MEANS |
*------------------------------*

%DO LI= 1 %TO &BYLEVEL;
%LET BEGROW =
%EVAL((&LI-1)*&VAR_NUM+1); %LET ENDROW =
%EVAL(&LI*&VAR_NUM);
XBETA = X*BETA[&LI, ];
XCovX =
VECDIAG(X*COVS[&BEGROW:&ENDROW,]*X)
;
%IF %UPCASE(&MODEL) = REG %THEN %DO;
CILOW=XBETA-1.96*SQRT(XCovX);
CIUP =XBETA+1.96*SQRT(XCovX);
ADJVALUE = XBETA;
%END;
%ELSE %IF %UPCASE(&MODEL)=LOGISTIC %THEN %DO;
LOW95=XBETA-1.96*SQRT(XCovX);
UP95 =XBETA+1.96*SQRT(XCovX);
CILOW=EXP(LOW95)/
(1+EXP(LOW95));
CIUP =EXP(UP95)/
(1+EXP(UP95));
ADJVALUE=EXP(XBETA)/
(1+EXP(XBETA));
%END;
CIMATRIX=INIT| CILOW | CIUP | ADJVALUE | XBETA;
CINAMES={&IX_VAR CILOW CIUP ADJVALUE XBETA};
CREATE CI&LI FROM IMATRIX[COLNAME =
CINAMES];
APPEND FROM CIMATRIX;
CLOSE CI&LI;
%END;
QUIT;

DATA TEMP;
LENGTH YNAME $8;
%IF %CASE(&MODEL) = LOGISTIC %THEN %DO;
IF _N_ = 1 THEN SET

YSUM(KEEP=YNUMBER); %END;
SET %DO QI=1 %TO &BYLEVEL;
CI&QI (IN=IN&QI)
%END; %IF %LENGTH(&BYVAR)>0 %THEN %DO;
%DO SI = 1 %TO &BY_NUM;
%LET BYV =
%SCAN(&BYVAR,&SI);
%DO QQ = 1 %TO &BYLEVEL;
IF IN&QQ THEN &BYV ="&BYV.&SI.&QQ";
%END;
%END;
%END;
%END;

YNAME =%UPPER("&YVAR");
RUN;

%IF &OUT= %THEN %DO;
%LET OUTDATA = OUT;
%END;
%ELSE %DO;
%LET OUTDATA =&OUT;
%END;
PROC APPEND BASE = &OUTDATA
DATA = TEMP;
%END;

%IF %UPCASE(&MODEL) = LOGISTIC %THEN %DO;
DO; PROC FREQ DATA = &INFILE
NOPRINT;
%IF %LENGTH(&BYVAR)>0 %THEN %DO;
TABLE &BYSORT*IXVAR/OUT=XNNUM;
%END;
%ELSE %DO;
TABLE &IXVAR/OUT= XNNUM;
%END;
RUN;
PROC SORT DATA = &OUTDATA;
BY &IX_VAR;
RUN;

DATA &OUTDATA;
MERGE &OUTDATA(IN=INA)
XNNUM(KEEP=BYVAR &IX_VAR
COUNT RENAME=(COUNT=XNUMBER));
BY &BYVAR &IX_VAR;
IF INA;
NPREDICT=XNUMBER*ADJVALUE;
RUN;
PROC SORT OUT=ALLSORT;
BY YNAME;
RUN;
PROC MEANS NOPRINT;
BY YNAME;
VAR NPREDICT;
OUTPUT OUT = OVERALL SUM
= OVERALL;
RUN;
PROC MEANS DATA = ALLSORT NOPRINT;
   BY YNAME;
   VAR YNUMBER;
   OUTPUT OUT = SUMY MEAN = SUMY;
RUN;

DATA &OUTDATA;
   MERGE ALLSORT
   SUMY (KEEP = YNAME SUMY)
   OVERALL (KEEP = YNAME OVERALL);
   BY YNAME;
   K = SUMY/OVERALL;
   ADJUSTED = K*ADJVALUE;
   LOWCI = K*CILow;
   UPCI = K*CUp;
RUN;
%END;

%IF %UPCASE(&MODEL) = REG %THEN %DO;
   DATA &OUTDATA;
   SET &OUTDATA;
   ADJUSTED = ADJVALUE;
RUN;
%END;
%IF %UPCASE(&PRINT) = YES %THEN %DO;
   PROC PRINT DATA = &OUTDATA;
RUN;
%END;
%IF %UPCASE(&MODEL) = LOGISTIC %THEN %DO;
   %LET TIT = PROPORTIONS;
   %LET TIT2 = P;
%END;
%ELSE %DO;
   %LET TIT = MEANS;
   %LET TIT2 = MEANS;
%END;

PROC TABULATE NOSEPS
ORDER = FORMATTED;
   CLASS YNAME &IX_VAR &BYVAR;
   VAR ADJUSTED CILOW CIUP;
   TABLE (YNAME = "" &IXTAB= ""),
   %IF %LENGTH(&BYVAR) > 0 %THEN
   %DO;
   &BYTAB = " ";
   (ADJUSTED = "ADJUSTED 
   &TIT2 = "SUM = " *F = 9.2
   CILOW = "LOWER 95% CI 
   &TIT2 = "SUM = " *F = 9.2
   CIUP = "UPPER 95% CI 
   &TIT2 = "SUM = " *F = 9.2
   /BOX = ". &UPCASE (&IX_VAR)
   RTS = TAB_NUM;
   TITLE "ADJUSTED &TIT AND ITS 95%
   CONFIDENCE INTERVALS";
   %END;
RUN;
%MEND ADJ_PROP;

Examples

Example 1. Adjusted Proportions

There is a SAS dataset called ONE in which included variables: Gender, Age, STIA, Race, Hypert, Income, Chol and Trig.

To calculate adjusted proportion of Hypert and Income with STIA after controlling Race and Age by Gender, call above the %ADJ_PROP macro with LOGISTIC option. Here the mean of age is equal to 55. The HYPERT, INCOME, RACE and STIA are 0 and 1 variables. After running macro the output table and a file called as OUTPROP can be obtained as below.

Example 2. Adjusted Means

Also one can use the ADJ_PROP macro to calculate adjusted means by replacing LOGISTIC by REG. Here, we consider the adjusted means of the continuous variable CHOL and TRIG after controlling age, race by gender.

Conclusion

Although there exists a lot paper concerning Logistic adjustment of proportion, it's very difficult to one who is unfamiliar with this method. Consequently, the macro ADJ_PROP presented here gives a useful and easy tool to calculate adjusted proportions and means.

Author Contact

Dingyi Zhao (uccdyz.cscc@mhs.unc.edu)
Dept. of Biostatistics,
University of North Carolina - CH Collaborative Studies Coordinating Center
137 E. Franklin Street, Suite #203
Chapel Hill, NC 27514
(919) 962-6971
Example 1: ADJUSTED PROPORTIONS AND ITS 95% CONFIDENCE INTERVALS

<table>
<thead>
<tr>
<th></th>
<th>STIA</th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ADJUSTED LOWER 95%</td>
<td>ADJUSTED LOWER 95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UPPER 95%</td>
<td>UPPER 95%</td>
</tr>
<tr>
<td>HYPERT</td>
<td>No</td>
<td>0.54</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.61</td>
<td>0.56</td>
</tr>
<tr>
<td>INCOME</td>
<td>No</td>
<td>0.22</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>0.13</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Example 2: ADJUSTED MEANS AND ITS 95% CONFIDENCE INTERVALS

<table>
<thead>
<tr>
<th></th>
<th>STIA</th>
<th>F</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ADJUSTED MEANS</td>
<td>ADJUSTED MEANS</td>
</tr>
<tr>
<td></td>
<td></td>
<td>LOWER 95%</td>
<td>LOWER 95%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>UPPER 95%</td>
<td>UPPER 95%</td>
</tr>
<tr>
<td>CHOL</td>
<td>No</td>
<td>218.60</td>
<td>209.85</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>224.32</td>
<td>214.07</td>
</tr>
<tr>
<td>TRIG</td>
<td>No</td>
<td>122.36</td>
<td>139.38</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>132.07</td>
<td>147.32</td>
</tr>
</tbody>
</table>

(6)