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:

Logit(P(Y=1|X)) = b0 + b1*x1 + ... + bixi + ... + bpxp;

Consider relationship between Y with xi after controlling the rest of x variables,where xi = 0 and 1 value .

Assume betas = (beta0 beta1 ... betap) is bs=(b0 b1 ... bp) estimators, respectively.

```
 MX = (1 mx1 ... 0 ... mxp,
       1 mx1 ... 1 ... mxp)
```

where mxj is mean of xj (j=1,...p and not equal i). Then, P at mean of X is

```
P(Y=1|MX) = exp(beats*MX')/
        (1+exp(betas*MX'));
```

and its 95% CI is

```
95%CI = exp(ci)/
        (1+exp(ci));
```

where **ci = betas*MX'+ 1.96*sqrt(VERDIAG(MX*COV*MX'));**

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)) = E(f(x)). So a corrected coefficient ,K, have to be used:

```
K=Actual Proportions overall/
predicted proportion overall;
 = P(Y=1)/
   ((p0 p1)*(P(Y=1|MX))');
```

where pi are the proportions of xi = 0 and 1, respectively.
The corrected proportion and its 95% CI are

```
Pc = K*P(Y=1|MX);
95%CIc = K*(95%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:

```
*_____________________________*
| Macro Name: ADJ_PROP
| VERSION   : 1
| FUNCTION  : MACRO ADJ_PROP
|             CAN BE USED TO
|             CALCULATE
|             ADJUSTED MEANS
|             AND PROPORTIONS.
|---------------------------
| 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,GIVENVAL,
|PRINT,DES,OUT)
|
|EXAMPLE USAGES:
|
|%ADJ_PROR(ONE,REG,GENDER,HDL,STIA,
|FB,AGE,55,YES,,RB).
|%ADJ_PROP(TWO,LOGISTIC,GENDER,CHD,
|STIA,FB,AGE,55,YES,DES,RW)
|-----------------------------
| 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,
GIVENVAL,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(&CONT_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 =
             %SCAN(&TVAR,&DI);
     %END;
     %LET TAB =&TAB&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&CONT_NUM;
 %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; END;
  %IF &GIVENVAL> .Z %THEN %DO;
      &GIVENVAR= &GIVENVAL;
  %END;
  INT = 1;
  RUN;

*---------------------------*
|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 &CONT_NUM;
     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; RESET PRINT;
  %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 PREDICATED       |
| 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 =BETA+1.96*SQRT(XCOVX);
 CILOW=EXP(LOW95)/
       (1+EXP(LOW95));
 CIUP =EXP(UP95)/
       (1+EXP(UP95));
 ADJVALUE=EXP(XBETA)/
          (1+EXP(XBETA));
%END;
  CIMATRX=INIT||CILOW||
          CIUP||ADJVALUE||XBETA;
  CINAMES={&IX_VAR CILOW CIUP
           ADJVALUE XBETA};
  CREATE CI&LI FROM IMATRX[COLNAME =
                          CINAMES];
  APPEND FROM CIMATRX;
  CLOSE CI&LI;
 %END;
 QUIT;

 DATA TEMP;
  LENGTH YNAME $8;
  %IF PCASE(&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
         ="&&BY&SI.&QQ";
       %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; 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 &BYVAR &IX_VAR;
RUN;

DATA &OUTDATA;
 MERGE &OUTDATA(IN=INA)
    XNNUM(KEEP=&BYVAR &IX_VAR
    COUNT RENAME=(COUNT=XNUMBER));
    BY &BYVAR &IX_VAR;
    IF INA;
    NPREDIT=XNUMBER*ADJVALUE;
RUN;

PROC SORT OUT=ALLSORT;
     BY YNAME;
RUN;

PROC MEANS NOPRINT;
     BY YNAME;
     VAR NPREDIT;
     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*CIUP;
 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=''*
   %END;
   (ADJUSTED="ADJUSTED
   &TIT2"*SUM=''*F=9.2
   CILOW='LOWER 95%CI'*SUM=''*F=9.2
   CIUP='UPPER 95%CI'*SUM=''*F=9.2)
   /BOX=".%UPCASE(&IX_VAR)"
    RTS=&TAB_NUM;
  TITLE "ADJUSTED &TIT AND ITS 95%
         CONFIDENCE INTERVALS";
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.

```
%let ys = HYPERT INCOME;
%adj_prop(one,logistic,gender,&ys,
stia,race,age,55,,des,outprop);
```

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.

```
%let ys = CHOL TRIG ;
%adj_prop(one,reg,gender,&ys,stia,
race,age,55,,,outmean);
```

Conclusion

Although there exists a lot paper concerning Logistic adjustment of proportion, it's very difficut 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 PROPORPTIONS AND ITS 95% CONFIDENCE INTERVALS

.	STIA	F			M		
		ADJUSTED P	LOWER 95% CI	UPPER 95% CI	ADJUSTED P	LOWER 95% CI	UPPER 95% CI
HYPERT	No	0.54	0.55	0.60	0.52	0.52	0.58
	Yes	0.61	0.53	0.75	0.56	0.44	0.74
INCOME	No	0.22	0.19	0.23	0.39	0.34	0.41
	Yes	0.13	0.07	0.22	0.23	0.11	0.40

Example 2: ADJUSTED MEANS AND ITS 95% CONFIDENCE INTERVALS

.	STIA	F			M		
		ADJUSTED MEANS	LOWER 95% CI	UPPER 95% CI	ADJUSTED MEANS	LOWER 95% CI	UPPER 95% CI
CHOL	No	218.60	217.48	219.72	209.85	208.71	210.99
	Yes	224.32	218.57	230.06	214.07	208.41	219.74
TRIG	No	122.36	120.28	124.44	139.38	136.57	142.19
	Yes	132.07	121.21	142.93	147.32	133.04	161.60