The SAS %RELIBPLS Macro

Roger Logan and Donna Spiegelman

July 26, 2004

Abstract

The macro %relibpls calculates regression coefficients, their standard errors, and odds ratios, when relevant, and 95% confidence intervals for a biologically meaningful difference specified by the user (the "increments"), which are all corrected for measurement error in one or more model covariates. Linear (proc reg), logistic (proc logistic), survival and conditional logistic (proc phreg) and mixed (proc mixed) models are implemented. A reliability study is required to empirically characterize the measurement error model. Details are given in Rosner et al. (1989), Rosner et al. (1990), and Rosner et al. (1992) including "real data" examples.

Keywords: SAS, macro, measurement error, reliability study

Contents

1 Description **1**

2 Invocation **2**

3 Examples **4**

4 Examples with output **5**

5 References **11**

6 Credits **11**

1 Description

This method of correction for logistic regression coefficients and related quantities is strictly valid only under the following conditions:

1. Outcome is rare (overall o served disease pro ability < 0.05).

2. All model covariates measured with error are continuous (perfectly measured covariates may be either categorical or continuous)

3. Measurement error is not severe (e.g., reliability coefficient > 0.5)

The sensitivity of the method to departures from assumptions #1 and #3 were carefully studied in rosner et al. (1989). The sensitivity of the method to departures from assumptions #1 and #3 have not been studied to date. The sensitivity of the method to departures from the normality assumption of the measurement error model were studied by Carroll et al. (JRSS-B, 1990), where it was found that the method is quite robust to departures from the normality assumption. **All datasets cannot contain any missing data.**

The method requires three sas datasets:

1. Names and increments for the variables in the model.

2. Main study. The uncorrected regression parameter estimates and variance-covariance matrix can be calculated using proc logistic, proc phreg, proc reg, or proc mixed. In the case that proc mixed is used, there will be two data sets: one for the estimated betas (`b`) and one for the estimated variance-covariance matrix (`varb`).

3. Reliability study.

2 Invocation

To call %relibpls, your program must know where to look for it. The most efficient way is to include the satement in your SAS program.

```
options mautosource sasautos='/usr/local/channing/sasautos';
```

You must load in a main study and validation study data set to be used in the regression and for determining the error model. A data set of weights, for use in calculating the odds ratios, must be defined. The weights for the covariates should be defined in the same order as those listed in the model statement used the the main study regression.

To obtain the uncorrected estimates of the model parameters you must run your regression using the main study, making an output data set using the covout option. This can be done using either PROC REG, PROC LOGISTIC, PROC PHREG, or PROC MIXED. For example using a logistic regression ,

```
proc logistic data=<main study data set name> covout outest=bvarbm;
  model <dependent variable> = <covariates> ;
```

This makes a (non-permanent) SAS data set (`bvarbm`) with the estimates and the variance-covariance matrix for the parameters.

You then type

```
%relibpls(data  =     <reliability study >,
          reg_ests =  <estimates of beta and cov(beta) from main study>,
          weights  =  < data set with weights used in odds ratio calcuations>,
          woe      =  < variables measured without error >,
          weri     =  <variables measured with error   >,
          rr       =  < number of replicates > ,
          varb     =  < cov(beta) when reg_type = MIXED . need two data sets, default = NONE >,
          reg_type =  < type of regression used. Can be LOGISTIC, PHREG, REG, MIXED,
                        or UNDETERMINED. Default is UNDETERMINED>,
          main     =  < Name of main study data set. Default =NONE>,
          out      =  <Name of data set for output. Default = NONE>
              );
```

Arguments to the macro are:

1. Name of the reliability dataset.

2. Name of the outest= dataset from the proc logistic, phreg, or reg with the covout option. Note that the independent variables must be in the same order as the names in the names and weights dataset specified in argument 3. In the case that proc mixed is used, this is the name of the data set containing the estimates of the uncorrected betas o tained using the `make 'solutionf' out=b;` option. In this case, the option `make 'covb' out=covbin ;` will need to be used for the input in argument 7 (see below).

3. Name of the data set containing the variable names and increments for the odds ratio or regression slopes. Note that the variables for which reliability measures have been taken must precede the variables which are perfectly measured.

4. Names of the variables in the estimate dataset specified in argument 2 which have been perfectly measured. The variables must be in the same order as the names in the names and weights dataset specified in argument 3.

5. Names of the variables in the estimate dataset specified in argument 2 which have reliability measures. These should have a set of variable names for each reliabilty measure (e.g. `x1 y1 z1 x2 y2 z2` for 2 measures). The variables must be in the same order as the names in the names and increments dataset specified in argument 3.

6. Number of reliabilty measures taken (e.g. 2).

7. (required if proc mixed is used; otherwise ignore) Name of data set containing covariance matrix of estimated beta in argument 2, if proc mixed was used in generating the estimates. The data set is o tained by using the option: `make 'covb' out=covbin`; the usage in the macro is `covb = covbin`.

8. (optional) Name of the main study dataset if availabe and used for calculating the estimates in argument 2. Usage is `main = mainin`.

9. (optional) SAS procedure used for calculating estimates used in argument 2. Can be one of logistic, phreg, reg, or mixed. If proc logistic, prog phreg, proc reg, or proc mixed was used in generating the estimates in argument 2 and 7. If the output from one of the above procs

was not used in argument 2, then this variable should not be set. This variable is used in determining the structure of the data sets containing the uncorrected estimates of beta and the covariance matrix of beta. The default value is `UNDETERMINED`. Usage is `reg_type=logistic`. This variable is also used in determining the type of regression used, logistic or linear. If this variable is not set, the default is taken to be logistic.

10. (optional) Name of output data set containing results of macro. Usage is `out=OUT`.

3 Examples

Let relib be the relibility study data set, `nm_wgt` the data set containing the increments (as described above), `age45,age50,age55,cigaret` the perfectly measured covariates and `cholest, glucose,bmi,sbp, cholest2, glucose2,bmi2,sbp2` the covariates measured with error. Let `rr=2` be the number of replicates.

1) No main study, where estimates of beta and variance are in a single data set (bvarb). The form of the data set should be

eta estimates covariance matrix of betas

```
%relibpls( data=relib,
           reg_ests=bvarb,
           weights=nm_wgt,
           woe=age45 age50 age55 cigaret,
           weri=cholest glucose bmi sbp cholest2 glucose2 bmi2 sbp2,
           rr=2);
```

2) Including a main study and using output from proc logistic, proc phreg, or proc reg. Let main e the main studywhich was used in the proc statement and bvarb the ouput of the estimates of eta and the covariance matrix, using the covout and outest= options. Suppose that we have used prog logistic.

```
proc logistic data=main covout outest=bvarb;
model case= cholest glucose bmi sbp age45 age50 age55 cigaret;
%relibpls( data=relib,
           reg_ests=bvarb,
           weights=nm_wgt,
           woe=age45 age50 age55 cigaret,
           weri=cholest glucose bmi sbp cholest2 glucose2 bmi2 sbp2,
           rr=2,
           reg_type = LOGISTIC
           main=main);
```

For linear regression using proc reg.

```
proc reg data=main covout outest=bvarb;
model case= cholest glucose bmi sbp age45 age50 age55 cigaret;
%relibpls( data=relib,
           reg_ests=bvarb,
           weights=nm_wgt,
           woe=age45 age50 age55 cigaret,
           weri=cholest glucose bmi sbp cholest2 glucose2 bmi2 sbp2,
           rr=2,
           reg_type = REG
           main=main);
```

3) Including a main study and using PROC MIXED. Suppose that we have used the proc mixed with the following commands. The data set `b` contains the estimates of beta and `covb` the estimates of the covariance matrix.

```
proc mixed method=ml data=main ;
model case= cholest glucose bmi sbp age45
             age50 age55 cigaret/ s covb;
make   'solutionF' out=b;
make   'covb'      out=covb;


%relibpls( data=relib,
           reg_ests=b,
           weights=nm_wgt,
           woe=age45 age50 age55 cigaret,
           weri=cholest glucose bmi sbp cholest2 glucose2 bmi2 sbp2,
           rr=2,
           varb = covb,
           reg_type = MIXED,
           main=main);
```

4 Examples with output

```
/* Examples for running the macro relibpls for the various
   types of regressions run. Can include logistic, reg, phreg,
   or mixed. */

%include 'relibpls.sas';

data test;
   input y x s case age;
   cards;
   1 2 2  0  10
   1 2.1 5 1 10
   2 3  4 0  11
```

```
   3 4 3  1  12
   2 2.9 3 0 13
   3 4.1 5 0 13
;

data relib;
input x x2 x3;
cards;
1.1 1.2 0.9
1.2 1.2 1.0
0.8 0.9 1.3
;

data wts;
input name $ weight;
cards;
x 1.0
s 1.0
;
```

```
title "proc reg";
proc reg covout outest=bvarb  data=test;
model y=x s ;

/* Example using linear regression */

%relibpls( data=relib,
           reg_ests=bvarb,
           weights=wts,
           woe=s,
           weri=x x2 x3,
           rr=3,
           reg_type=reg,
           main=test);

OUTPUT:

CORRECTION OF RELATIVE RISK ESTIMATES FROM REG REGRESSION FOR RANDOM WITHIN-PERSON MEASUREMENT ERROR

Reference: Rosner BA, Spiegelman D, Willett WC. 'Correction of  ogistic Regression Relative Risk
           Estimates and Confident Intervals for Random Within-Person Measurement Error'
            Amer. J. Epi, 1992;136: 1400-1413 .
Programmers:  Albert  iou, M.S., The Channing  aboratory, Boston, MA.
              Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston, MA.
              Carrie Wager, B.S., The Channing  aboratory, Boston, MA.

Random Within-Person Covariance Matrix:

                  x

x          .556E-02


Random Between-Person Covariance Matrix:

                  x          s

x         7.701E-01  2.667E-01
s         2.667E-01  1.467E+00


REG Regression:

             **************Uncorrected***************    *************Corrected*************
               WT       B      SE(B)       95% CI(B)        B      SE(B)      95% CI(B)

x            1.000    1.009    0.034    0.942 - 1.075     1.058    0.056    0.949 - 1.167
s            1.000   -0.047    0.025   -0.096 - 0.002    -0.056    0.031   -0.116 - 0.004
```

```
/* Example using logistic regression */

title "proc logistic";
proc logistic descending covout outest=bvarb data=test noprint;
model case = x s;
proc print data=bvarb;

%relibpls( data=relib,
           reg_ests=bvarb,
           weights=wts,
           woe=s,
           weri=x x2 x3,
           rr=3,
           reg_type=logistic,
           main=test);

OUTPUT:

CORRECTION OF RELATIVE RISK ESTIMATES FROM  OGISTIC REGRESSION FOR RANDOM WITHIN-PERSON MEASUREMENT
ERROR

Reference: Rosner BA, Spiegelman D, Willett WC. 'Correction of  ogistic Regression Relative Risk
           Estimates and  Confident Intervals for Random Within-Person Measurement Error'
            Amer. J. Epi, 1992;136: 1400-1413 .
Programmers:  Albert  iou, M.S., The Channing  aboratory, Boston, MA.
              Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston, MA.
              Carrie Wager, B.S., The Channing  aboratory, Boston, MA.


Random Within-Person Covariance Matrix:

                  x

x          .556E-02

Random Between-Person Covariance Matrix:

                  x          s

x         7.701E-01  2.667E-01
s         2.667E-01  1.467E+00


 OGISTIC Regression:

               **************Uncorrected**************       ***************Corrected***************
              WT       B      SE(B)    OR(B)      95% CI(OR)        B      SE(B)    OR(B)      95%
CI(OR)

x           1.000   -0.048    1.103    0.95     0.110 - 8.281    -0.050    1.157    0.951    0.098 - 9.190
s           1.000    0.430    0.8 5    1.537    0.299 - 7.890     0.430    0.836    1.5 8    0.299 - 7.922
```

```
/* Example using Cox's proportional hazards regression */

title "proc phreg";
proc phreg data=test covout outest=bvarb noprint;
model age*case(0)= x s / ties=breslow;

%relibpls( data=relib,
           reg_ests=bvarb,
           weights=wts,
           woe=s,
           weri=x x2 x3,
           rr=3,
           reg_type=phreg,
           main=test);

OUTPUT:

proc phreg

CORRECTION OF RELATIVE RISK ESTIMATES FROM COX REGRESSION FOR RANDOM WITHIN-PERSON MEASUREMENT ERROR

Reference: Rosner BA, Spiegelman D, Willett WC. 'Correction of  ogistic Regression Relative Risk
           Estimates and Confident Intervals for Random Within-Person Measurement Error'
            Amer. J. Epi, 1992;136: 1400-1413.
Programmers:  Albert  iou, M.S., The Channing  aboratory, Boston, MA.
              Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston, MA.
              Carrie Wager, B.S., The Channing  aboratory, Boston, MA.


Random Within-Person Covariance Matrix:

                 x

x          .556E-02


Random Between-Person Covariance Matrix:

                 x          s

x         7.701E-01  2.667E-01
s         2.667E-01  1.467E+00


COX Regression:

                  ***************Uncorrected***************     ***************Corrected****************
              WT       B      SE(B)    OR(B)      95% CI(OR)        B      SE(B)    OR(B)      95%
CI(OR)

x           1.000   -0.874    1.165    0.417    0.043 - 4.095    -0.917    1.223    0.400    0.036- 4.395
s           1.000    0.482    0.712    1.619    0.401 - 6.5 8     0.490    0.715    1.632    0.401- 6.633
```

```
/* Example using mixed regression */

title "proc mixed";

proc mixed method=ml  data=test ;
model y = x s / s covb ;
   make 'covb' out=covb;
   make 'SolutionF' out=b;

%relibpls( data=relib,
           reg_ests=b,
           weights=wts,
           woe=s,
           weri=x x2 x3,
           rr=3,
           varb=covb,
           reg_type=mixed,
           main=test);
```

OUTPUT:

CORRECTION OF RELATIVE RISK ESTIMATES FROM MIXED REGRESSION FOR RANDOM WITHIN-PERSON MEASUREMENT ERROR

Reference: Rosner BA, Spiegelman D, Willett WC. 'Correction of ogistic Regression Relative Risk Estimates and Confident Intervals for Random Within-Person Measurement Error' Amer. J. Epi, 1992;136: 1400-1413.

Programmers: Albert iou, M.S., The Channing aboratory, Boston, MA.
Donna Spiegelman, Sc.D., Harvard School of Public Health, Boston, MA.
Carrie Wager, B.S., The Channing aboratory, Boston, MA.

Random Within-Person Covariance Matrix:

	x
x	.556E-02

Random Between-Person Covariance Matrix:

	x	s
x	7.701E-01	2.667E-01
s	2.667E-01	1.467E+00

MIXED Regression:

	**********Uncorrected**********				**********Corrected**********		
	WT	B	SE(B)	95% CI(B)	B	SE(B)	95% CI(B)
x	1.000	1.009	0.024	0.962 - 1.055	1.058	0.050	0.961 - 1.155
s	1.000	-0.047	0.018	-0.082 - -0.012	-0.056	0.025	-0.105 - -0.007

5 References

Rosner, Spiegelman, Willett. Correction of logistic regression relative risk estimates and confidence intervals for random withinperson measurement error. American Journal of Epidemiology 1992; 136: 1400-1413.

6 Credits

Programmers: Donna Spiegelman, Albert Liou, Carrie Wager, Marcia Goetsch, Roger Logan, Ruifeng Li, for the Channing Laboratory.

For further information contact: Dr. Donna Spiegelman, `stdls@channing.harvard.edu`, (617) 432-0205.