Using the SYMPUT Function to Automatically Choose Reference for Bivariate Cox Proportional Models

Zhongjie Cai, University of Southern California; Li Ding, University of Southern California

ABSTRACT

Bivariate Cox Proportional Models are used when you test the association between a single covariate and the outcome for survival analysis. This procedure repeats for each covariate of interest. As SAS uses the last category as the default reference in CLASS statement, you will encounter problems when you would like to use other categories as references. If a format is associated with that covariate, you will have to use the corresponding format instead of the raw numeric data. This problem becomes even worse when you have to repeat the procedure and hand type the reference every single time. This presentation demonstrates one way of fixing the problem using SAS MACRO and the CALL SYMPUT function.

INTRODUCTION

The main goal of this MACRO is to reduce human effort. Rather than manually typing the formatted reference categories for each covariate, you can use the CALL SYMPUT function to assign the reference format to a MACRO variable.

This paper focuses on Proportional Hazards model (the PHREG procedure) but can be easily modified to fit in other models (e.g. the LOGISTIC procedure) as long as a CLASS statement is available.

DATA STRUCTURE AND FORMAT

For the purpose of demonstration, let's use the BMT data set in the SASHELP library. And to demonstrate formatting, an arbitrary Age Group variable is added with format. Below is the variable list of the dataset (Table 1) obtained by the CONTENTS procedure and the corresponding format created (Table 2):

Alphabetic List of Variables and Attributes				
#	Variable	Type	Len	Format	Label
1	Group	Char	13		Disease Group
3	Status	Num	8		Event Indictor: 1=Event 0=Censored
2	T	Num	8		Disease-Free Survival Time
4	age	Num	8		
5	agegroup	Num	8	AGEGROUPF.	

Table 1. Data Description from the CONTENT procedure

```
                    FORMAT NAME: AGEGROUPF LENGTH: 5
    MIN LENGTH:   1   MAX LENGTH:   40  DEFAULT LENGTH:   5  FUZZ: STD

START                  END                  LABEL  (VER. 9.4      02FEB2017:20:31:30)

                    0                   0 <5
                    1                   1 5-10
                    2                   2 10-15
                    3                   3 >15
```

Table 2. Data Format

Assuming AGEGROUP is the variable you would like to look at, if you use REF=FIRST then you would get age group 10-15 as the reference (Output 1. Output of Class Level Information with REF=FIRST).

Class Level Information				
Class	**Value**	**Design Variables**		
agegroup	10-15	0	0	0
	5-10	1	0	0
	<5	0	1	0
	>15	0	0	1

Output 1. Output of Class Level Information with REF=FIRST

To change the reference level to "<5", you will need to use REF="<5" in the CLASS statement. This reference will have to be specified for every single formatted covariate.

MACRO

The MACRO uses the CALL SYMPUT function to assign the format label corresponding to 0 ("<5" in previous example). The code is shown below:

```
%macro uni(var,time,event,ref=0);
      /*
      var: covariate of interest
      time: time to event
      event: event indicator, 0=Censored 1=Event
      ref: numeric reference level, default 0
      */
      data _NULL_;
            call symput('varformat',put(&ref,&var.f.));
            *assign the format corresponding to "ref" to global variable
"varformat", default=0;
      run;
      proc phreg data=bmt;
            class &var (param=ref ref="&varformat");
            *must use double quote to do substitution, not single;
            model &time*&event(1)=&var;
      run;
   %mend;
```

The MACRO assumes the variable's format is named after itself with an additional f at the end. For example the format associated with variable AGEGROUP should be AGEGROUPF (modifiable).

Below are sample outputs with different reference level specified (Output 2 and Output 3):

Class Level Information				
Class	Value	Design Variables		
agegroup	10-15	1	0	0
	5-10	0	0	0
	<5	0	1	0
	>15	0	0	1

Output 2. Partial Output from %uni(agegroup,t,status,ref=1)

Class Level Information				
Class	Value	Design Variables		
agegroup	10-15	0	0	0
	5-10	1	0	0
	<5	0	1	0
	>15	0	0	1

Output 3. Partial Output from %uni(agegroup,t,status,ref=2)

CONCLUSION

Bivariate Cox Proportional Models are widely used at an early stage of analyzing survival data. This paper has shared a simple MACRO that helps to reduce manual labor when you have to specify different reference levels.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Terry Zhongjie Cai
University of Southern California
terry.zhongjiecai@gmail.com

Li Ding
University of Southern California
dingli.pumc@gmail.com