**Conduct Backward Selection in Multivariable Survival Analysis**

**Macro:** PHREG\_SEL

**Created Date/Author/Contact:** June 20, 2012/Dana Nickleach

**Last Update Date/Person/Contact**: Oct 2016/Yuan Liu

**Other Significant Contributor**: Mar 2016/Yaqi Jia

**Current Version**: V23

**Working Environment:** SAS 9.4 English version

**Contact**: Dr. Yuan Liu [yliu31@emory.edu](mailto:yliu31@emory.edu)

**Purpose:** To conduct backward selection on a Cox proportional Hazard model using the maximum possible sample size at each stage of the selection process instead of restricting to the sample size from the first step as SAS does when using their selection methods. Optionally, a table of the resulting model can be generated.

**Notes:** The model is run using PROC PHREG. The final list of variables selected will be written to the log. Additionally, two global macro variables, \_finalvar and \_finalcvar will be created containing the list of all variables and categorical variables selected, respectively. If you are requesting a table with the model results then the macro “MUTLIPLE\_PHREG V21” or later is also required. Interactions can be included to obtain the estimate of treatment effect (TRT) in each level of stratified variable (SV), and it is required both TRT and SV to be categorical variables. For variables selection, put TRT, SV, and TRT\*SV in the beginning of VAR; use INC = 3 to force the two main effects and their interaction in the model; use EFFECT = TRT and SLICEBY = SV to generate the stratified treatment hazard ratio.

**Parameters:**

|  |  |
| --- | --- |
| **Macro variable** | **Description** |
| DSN | The name of the data set to be analyzed. |
| EVENT | Time to event outcome variable. |
| START | Instead of a single failure-time variable (EVENT) a pair of failure time variables can be specified. This should be the start point of the interval for which the subject is at risk. The label of this variable will be used in the table header. |
| STOP | Instead of a single failure-time variable (EVENT) a pair of failure time variables can be specified. This should be the end point of the interval for which the subject is at risk. |
| CENSOR | Name of censoring indicator variable. Values of 0 indicate censored observations. |
| VAR | List of variables to include in the model separated by spaces. |
| CVAR | The list of categorical variables that are in VAR. If need to change the reference group, you can follow each variable name by (DESC) or by (ref = “Ref level in formatted value”) where needed and separate terms by \*. See code example. |
| INC | Number of variables to include in the model (optional). The first *n* variables in the VAR parameter will be forced in the model. The default value is 0. |
| SLSTAY | The significance level for removing variables from the model (optional). The default value is .05. |
| WEIGHT | Variable to use in the weight statement (optional). Leave it blank if not using weights. The number of observations reported in the table footer are not weighted. |
| WEIGHTNORM | Set this to F if you don't want the weights to be normalized to the original sample size using the normalize option (optional). The default value is T. |
| ID | ID variable for identifying observations to be used in the ID statement in PHREG (optional). If observations have more than one record in the data file, the number of observations reported in the table footer will be the number of records, not unique observations. |
| COVSAGG | Set to T to use the COVS(AGGREGATE) option in PHREG (optional). The default value is F. Be aware that backward selection does not use the COVS estimates as previously documented here: http://support.sas.com/kb/14/813.html. |
| STRATA | List of variables to stratify on separated by spaces (optional). These variables are not candidates for removal from the model. Leave it blank if not using strata. |
| REPORT | Set this to T if you want a table of the resulting model generated (optional). The default value is F. |
| TYPE3 | Set to F to suppress type III p-values from being reported in the table (optional). The default value is T. This only has an effect if REPORT = T. |
| EFFECT | Use to specify the treatment variable in the interaction. Use in combine with SLICEBY and if not empty, VAR should contain a two-way interaction. See example. |
| SLICEBY | Use in combine with EFFECT to specify the stratified variable in the interaction. |
| CLNUM | Set to T if you want to see the number of observations for each level of covariates. The default is F. |
| COMPARE | Use in combine with EFFECT and SLICEBY when there is an interaction in the model. Take value as CONTROL or ALL. ALL means to conduct pairwise comparison and CONTROL means only to conduct comparison with reference group. The default is CONTRL. |
| ORIENTATION | Orientation of the output Word table. Default is portrait, can be changed to landscape. |
| SHORTREPORT | Use in combine with EFFECT and SLICEBY when there is an interaction in the model and set to T to only report the stratified treatment effect. |
| FILENAME | File name for output table. This is necessary if report=T. |
| OUTPATH | File path for output table to be stored. This is necessary if report=T. |
| DEBUG | Set to T if running in debug mode (optional). Work datasets will not be deleted in debug mode. This is useful if you are editing the code or want to further manipulate the resulting data sets. The default value is F. |

**Usage Example:**

**DATA** analysis;

input id os\_censor Sex $ Age duration os progress $ trt $;

LABEL os = 'Overall Survival (months)'

progress = 'Progression'

trt = 'Treatment'

duration = 'Duration of Radiation';

DATALINES;

1 1 M 40 44 20 No B

2 1 F 45 46 16 Yes A

3 1 F 40 32 20 No B

4 1 F 47 32 23 No B

5 0 M 41 25 22 No B

6 1 M 54 35 13 No B

7 1 M 48 50 9 Yes A

8 1 M 36 33 12 Yes B

9 0 F 49 51 8 Yes A

10 1 M 49 52 10 Yes A

11 1 M 44 35 12 No A

12 1 M 49 50 8 Yes A

13 1 M 44 44 14 Yes A

14 1 M 50 31 10 Yes A

15 1 M 53 40 15 No B

16 0 M 52 29 20 Yes B

17 1 F 46 45 5 Yes A

18 1 F 37 44 11 Yes A

19 1 M 49 46 13 No B

20 1 M 42 31 11 No A

;

TITLE 'Table 4 Multivariable Survival Analysis';

%PHREG\_SEL(dsn = analysis,

event = os,

censor = os\_censor,

var = trt sex age duration,

cvar = trt(ref="A")\*sex,

inc = 1,

slstay = .20,

report = T,

outpath = C:\Documents and Settings\User\My Documents\,

filename = Table 4 Multivariable Survival Analysis);

TITLE;

TITLE 'Table 4-2 Multivariable Survival Analysis - stratified by Sex';

%PHREG\_SEL(dsn = analysis,

event = os,

censor = os\_censor,

var = trt sex trt\*sex ,

cvar = trt(ref="A")\*sex(DESC),

effect = trt,

sliceby = sex,

inc = 3,

slstay = .20,

clnum=T,

report = T,

outpath = C:\Documents and Settings\User\My Documents\,

filename = Table 4-2 Multivariable Survival Analysis - stratified by Sex);

TITLE;

This message will be written in the log:

Categorical variables selected: TRT SEX

All variables selected: SEX

**Summary Table Example:**

Table 4 Multivariable Survival Analysis

|  | | **Overall Survival (months)** | | |
| --- | --- | --- | --- | --- |
|  | | **----------------------------------------** | | |
| **Covariate** | **Level** | **Hazard Ratio (95% CI)** | **HR P-value** | **Type3 P-value** |
| Treatment | B | 0.20 (0.05-0.75) | **0.017** | **0.017** |
| A | - | - |
|  | | | | |
| Sex | F | 0.44 (0.13-1.50) | 0.189 | 0.189 |
| M | - | - |
|  | | | | |
| Duration of Radiation |  | 1.05 (0.98-1.14) | 0.181 | 0.181 |
|  | | | | |
| \*  Number of observations in the original data set = 20. Number of observations used = 20. \*\* Backward selection with an alpha level of removal of .20 was used. The following variables were removed from the model: Age. | | | | |

Table 4-2 Multivariable Survival Analysis - stratified by Sex

|  | | | **Overall Survival (months)** | | |
| --- | --- | --- | --- | --- | --- |
|  | | | **----------------------------------------** | | |
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Type3 P-value** |
| **Comparisons Stratified by Sex** | **Treatment :** |  | - | - | 0.874 |
|  | | | | | |
| M | A vs. B | 7 vs. 7 | 6.75 (1.70-26.82) | **0.007** | - |
|  | | | | | |
| F | A vs. B | 4 vs. 2 | 8.31 (0.76-90.69) | 0.082 | - |
|  | | | | | |
| \*  Number of observations in the original data set = 20. Number of observations used = 20. \*\* Backward selection with an alpha level of removal of .20 was used.  No variables were removed from the model. \*\*\* The estimated stratified treatement effect was controlled by: None | | | | | |

**Permission:**

Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the "Software"), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:   
  
The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.  
  
THE SOFTWARE IS PROVIDED "AS IS", WITHOUT WARRANTY OF ANY KIND, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO THE WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE AND NONINFRINGEMENT. IN NO EVENT SHALL THE AUTHORS OR COPYRIGHT HOLDERS BE LIABLE FOR ANY CLAIM, DAMAGES OR OTHER LIABILITY, WHETHER IN AN ACTION OF CONTRACT, TORT OR OTHERWISE, ARISING FROM, OUT OF OR IN CONNECTION WITH THE SOFTWARE OR THE USE OR OTHER DEALINGS IN THE SOFTWARE.

**Log of Updates:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **By** | **Description** | **Version** |
| 7/6/12 | Dana Nickleach (dnickle@emory.edu) | Added nowarn option to ODS output, added event and censor parameters, and set the inc parameter to a default value of 0. | V1.2 |
| 7/10/12 | Dana Nickleach (dnickle@emory.edu) | Added slstay parameter. | V1.2 |
| 7/31/12 | Dana Nickleach (dnickle@emory.edu) | Added strata parameter. | V2 |
| 8/6/12 | Dana Nickleach (dnickle@emory.edu) | Added the option of generating a table with model results. | V3 |
| 8/28/12 | Dana Nickleach (dnickle@emory.edu) | Added initial variable list to the footnote. | V4 |
| 9/19/12 | Dana Nickleach (dnickle@emory.edu) | Added check for variable names that are too long, modified footnote to only list removed variables, and fixed problem so that final list of selected variable does not get overwritten when calling multiple\_phreg. | V5 |
| 10/23/12 | Dana Nickleach (dnickle@emory.edu) | Prevented case sensitivity of T/F parameters. | V6 |
| 12/19/12 | Dana Nickleach (dnickle@emory.edu) | Fixed problem in update of variable list after removal and prevented ODS request for type III table when there are no categorical variables. | V7 |
| 1/4/13 | Dana Nickleach (dnickle@emory.edu) | Made changes to operate with the new version of multiple\_phreg and added the parameter TYPE3. | V8 |
| 1/22/13 | Dana Nickleach (dnickle@emory.edu) | Added handling of interaction terms for selection procedure. | V9 |
| 3/8/13 | Dana Nickleach (dnickle@emory.edu) | Added footnote about strata. | V10 |
| 5/3/13 | Dana Nickleach (dnickle@emory.edu) | Added DEBUG parameter and modified the call to MULTIPLE\_PHREG to be consistent with the new version. | V11 |
| 7/11/13 | Dana Nickleach (dnickle@emory.edu) | Added creation of global macro variables holding final variable list. | V12 |
| 7/16/13 | Dana Nickleach (dnickle@emory.edu) | Added option to set reference categories. | V13 |
| 7/23/13 | Dana Nickleach (dnickle@emory.edu) | Fixed case when only one class variable is specified along with (DESC). | V14 |
| 8/29/13 | Dana Nickleach (dnickle@emory.edu) | Declared macro variables as local. | V15 |
| 11/19/13 | Dana Nickleach (dnickle@emory.edu) | Modified boolean check for text in footnote2 to prevent bug. | V16 |
| 1/30/14 | Dana Nickleach (dnickle@emory.edu) | Modified so that there are not restrictions on the length of variable names other than the SAS allowances. | V17 |
| 2/12/14 | Dana Nickleach (dnickle@emory.edu) | Fixed so that DEBUG option carries through to MULTIPLE\_PHREG as well. | V18 |
| 2/13/14 | Dana Nickleach (dnickle@emory.edu) | Added the START, STOP, WEIGHTNORM, ID, and COVSAGG parameters. | V19 |
| 3/26/14 | Dana Nickleach (dnickle@emory.edu) | Fixed bug occurring when variables names are long that prevents variables from being dropped from the categorical variable list. | V20 |
| 2/4/2015 | Yuan Liu | Allow interaction in the model and estimated stratified treatment effect; add EFFECT, SLICEBY, SHORTREPORT macro variables | V21 |
| 10/16/15 | Yuan Liu | Add COMPARE parameter | V22 |
| 10/27/2016 | Yaqi Jian | Allow the number of observations for each variable to show in the final report. CLNUM was added | V23 |
| Oct, 2016 | Yuan Liu | Allow user to specify any reference level of a categorical variable. | V23 |