**Conduct Backward Selection in Multivariable Logistic Regression**

**Macro:** LOGREG\_SEL

**Created Date/Author** Oct. 19, 2012/Dana Nickleach

**Last Update Date/Person** Jan, 2017/Yuan Liu

**Other Significant Contributor:** Yaqi Jia

**Current Version**: V15

**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 logistic regression 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 LOGISTIC. A binary outcome or ordinal outcome using a cumulative logit model can be used, but not a nominal outcome. 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\_LOGREG V16” 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 odds ratio.

**Parameters:**

|  |  |
| --- | --- |
| **Macro variable** | **Description** |
| DSN | The name of the data set to be analyzed. |
| OUTCOME | The name of the outcome variable. It must be binary or ordinal. |
| EVENT | The event category for the binary response model. Specify the value in quotes. This is the argument that will be passed to the event= option in the model statement. Leave this blank if you have an ordinal outcome with more than 2 levels. |
| DESC | Set to T to reverse the order of an ordinal outcome (optional). The order will be based on the internal order. Only specify this if the EVENT parameter is blank. The default value is F. |
| VAR | List of variables to include in the model separated by spaces. |
| CVAR | The list of categorical variables that are in VAR, separated by spaces. 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 included in every 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). Weights will be normalized to the original sample size using the normalize option. Leave it blank if not using weights. |
| 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. |
| 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 T. |
| 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 Logistic Regression';

%logreg\_sel(dsn = analysis,

outcome = progress,

event = 'Yes',

var = Trt Sex Age Duration,

cvar = Trt(ref=”A”)\*Sex,

inc = 1,

slstay = .50,

report = T,

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

filename = Table 4 Logistic Regression);

TITLE;

TITLE 'Table 4-2 Logistic Regression – Stratify by SEX';

%logreg\_sel(dsn = analysis,

outcome = progress,

event = 'Yes',

var = Trt Sex Trt\*Sex Age Duration,

cvar = Trt(ref=”A”)\*Sex,

inc = 3,

effect = Trt,

sliceby = Sex,

slstay = .50,

report = T,

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

filename = Table 4-2 Logistic Regression – Stratify by SEX);

TITLE;

**Summary Table Example:**

Table 4 Logistic Regression

|  | | | **Progression=Yes** | | |
| --- | --- | --- | --- | --- | --- |
|  | | | **----------------------------------------** | | |
| **Covariate** | **Level** | **N** | **Odds Ratio (95% CI)** | **OR P-value** | **Type3 P-value** |
| Treatment | B | 9 | 0.10 (0.01-1.07) | 0.057 | 0.057 |
| A | 11 | - | - |
|  | | | | | |
| Duration of Radiation |  | 20 | 1.08 (0.93-1.27) | 0.318 | 0.318 |
|  | | | | | |
| \*  Number of observations in the original data set = 20. Number of observations used = 20. \*\* Backward selection with an alpha level of removal of .5 was used. The following variables were removed from the model: Age, and Sex. | | | | | |

Table 4-2 Logistic Regression – Stratify by SEX

|  | | | **Progression=Yes** | | |
| --- | --- | --- | --- | --- | --- |
|  | | | **----------------------------------------** | | |
| **Covariate** | **Level** | **N** | **Odds Ratio (95% CI)** | **OR P-value** | **Type3 P-value** |
| **Comparisons Stratified by Sex :** | **Treatment :** |  |  | - | 0.931 |
|  | | | | | |
| F | A vs. B | 4 vs. 2 | 5.5576E9 (0.00-1.14E211) | 0.924 | - |
|  | | | | | |
| M | A vs. B | 7 vs. 7 | 6.25 (0.61-63.54) | 0.121 | - |
|  | | | | | |
| \*  Number of observations in the original data set = 20. Number of observations used = 20. \*\* Backward selection with an alpha level of removal of .50 was used. The following variables were removed from the model: Age, and Duration of Radiation. \*\*\* The estimated stratified treatement effect was controlled by: &conlab | | | | | |

**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** |
| 10/19/12 | Dana Nickleach | Prevented case sensitivity of report parameter and modified counting of variables to only use spaces as a delimiter. | V2 |
| 11/2/12 | Dana Nickleach | Changed macro variable names holding final selection to be more unique in case the user needs to make them global. | V3 |
| 11/5/12 | Dana Nickleach | Added debug parameter. | V3 |
| 2/22/13 | Dana Nickleach | Updated to correspond with updates in multiple\_logreg V6. | V4 |
| 3/17/13 | Dana Nickleach | Added TYPE3 parameter. | V5 |
| 4/18/13 | Dana Nickleach | Fixed problem in update of variable list after variable removal and modified EVENT parameter so that it can also be used in the table header. | V6 |
| 5/3/13 | Dana Nickleach | Updated the call to MULTIPLE\_LOGREG based on the new version and fixed so that it will run without errors if there are no class variables. | V7 |
| 7/12/13 | Dana Nickleach | Updated the call to MULTIPLE\_LOGREG based on the new version, added creation of global macro variables holding final variable list, and set up to handle interaction terms in the selection process, not report. | V8 |
| 7/22/13 | Dana Nickleach | Added DESC parameter, option to change reference levels in CVAR, and added handling of ordinal outcomes. | V9 |
| 9/5/13 | Dana Nickleach | Declared macro variables as local. | V10 |
| 10/10/13 | Dana Nickleach | Fixed so that commas in the event label will not cause errors. | V11 |
| 2/6/14 | Dana Nickleach | Modified so that there are not restrictions on the length of variable names other than the SAS allowances. | V12 |
| 2/4/2015 | Yuan Liu | Allow interaction in the model and estimated stratified treatment effect; add EFFECT, SLICEBY, SHORTREPORT macro variables | V13 |
| Mar, 2016 | Yaqi Jia | Allow the number of observations for each variable to show in the final report. CLNUM was added | V14 |
| Jan, 2017 | Yuan Liu | Allow user to specify any reference level of a categorical variable. | V15 |