**Conduct Backward Selection in Multivariable General Linear Regression**

**Macro:** LINREG\_SEL

**Created Date/Author:** Oct. 3, 2013/Dana Nickleach

**Last Update Date/Person**:

**Current Version**: V1

**Working Environment:** SAS 9.3 English version

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

**Purpose:** To conduct backward selection on a general linear 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 GLMSELECT and PROC GLM. 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\_LINREG V2” or later is also required. Interactions can be included if you are not requesting a table (REPORT=F). Interactions will be removed from the model with respect to model hierarchy. However, the macro cannot produce a correct report table if the model contains interactions.

**Parameters:**

|  |  |  |
| --- | --- | --- |
| **Macro variable** | **Description** | **Required** |
| DSN | The name of the data set to be analyzed. | Yes |
| OUTCOME | The name of the continuous outcome variable. | Yes |
| VAR | List of all variables to include in the model separated by spaces. Each variable name or interaction term must not be more than 20 characters long. | Yes |
| CVAR | List of categorical variables to include in the model separated by spaces. These should also appear in the VAR parameter. | Yes |
| ORDER | The order in which to sort the levels of the categorical variables: DATA, FORMATTED, FREQ, or INTERNAL. The default value is INTERNAL. | No |
| INC | Number of variables to include in the model. The first *n* variables in the var parameter will be included in every model. The default value is 0. | No |
| SLSTAY | The significance level for removing variables from the model. The default value is .05. | No |
| WEIGHT | Variable to use in the weight statement. The macro will **not** normalize weights to the original sample size. Weights will need to be normalized in the data preparation stage. Leave it blank if not using weights. | No |
| REPORT | Set this to T if you want a table of the resulting model generated. The default value is F. | No |
| TYPE3 | Set to F to suppress type III p-values from being reported in the table. The default value is T. This only has an effect if REPORT = T. | No |
| OUTPATH | File path for output table to be stored. | Yes, if REPORT=T |
| FILENAME | File name for output table. | Yes, if REPORT=T |
| DEBUG | Set to T if running in debug mode. 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. | No |

**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 Multivariate Regression';

%linreg\_sel(dsn=analysis,

outcome=duration,

var=sex trt age,

cvar=sex trt,

inc=1,

slstay=.20,

report=T,

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

filename=Table 4 Multivariate Regression);

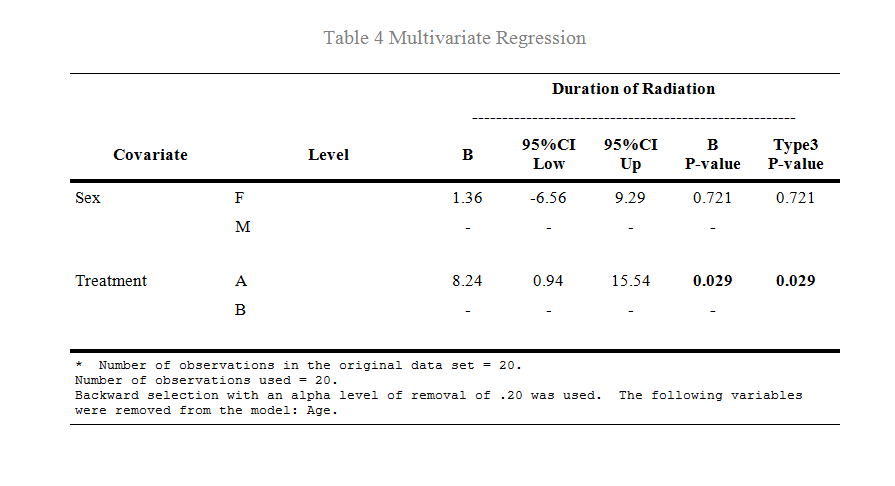
TITLE;

This message will be written in the log:

Categorical variables selected: SEX TRT

All variables selected: SEX TRT

**Summary Table Example:**



**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** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |