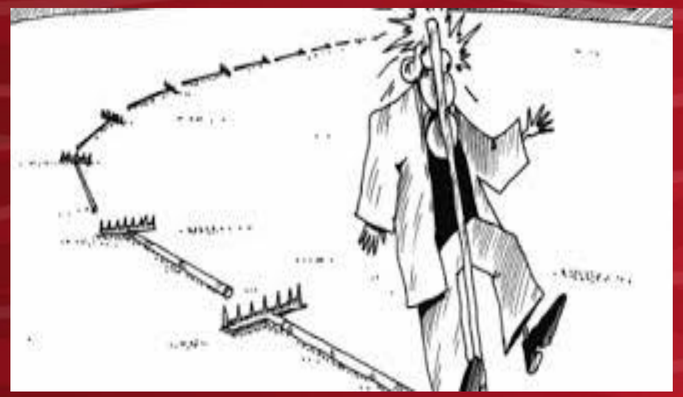
SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL

You Do Not Have To Step On The Same Rake: SAS Raking Macro – Generation IV



David Izrael, Abt Associates
Michael P. Battaglia, Battaglia Consulting Group, LLC.
Annabella A. Battaglia, Battaglia Consulting Group, LLC.
Sarah W. Ball, Abt Associates

USERS PROGRAM

You Do Not Have To Step On The Same Rake: SAS Raking Macro – Generation IV

David Izrael, Michael P. Battaglia, Ann A. Battaglia, Sarah W. Ball

ABSTRACT

In this poster, we present the 4th generation of the raking macro where we introduce a few helpful enhancements including: **a**) an explicit indicator for trimming (or not trimming) the weight which substantially saves run time when no trimming is needed; **b**) two methods of weight trimming – 'AND' and 'OR' – that enable users to overcome a stubborn non-convergence; **c**) the summary statistics related to the number of cases with trimmed or increased weights has been expanded; and **d**) we introduce parameters that enable users to use different criteria of convergence for different raking marginal variables.

RAKING MACRO: GENERATION IV

Macro parameters of the new version is shown on the right. All the new parameters relate to the weight trimming component of the macro.

a) An explicit indicator for trimming (or not trimming) the weight which substantially saves run time when no trimming is needed (**TRIMWEIGHT**) .

b) Two methods of weight trimming – 'AND' and 'OR' – that enable users to overcome a stubborn non-convergence (**METHOD, GL_SWITCH, GH_SWITCH, IL_SWITCH, IH_SWITCH, GL_AND_IL_SWITCH, GH_AND_IH_SWITCH**) .

c) The summary statistics related to the number of cases with trimmed or increased weights has been expanded.

d) We introduce parameters that enable users to use different criteria of convergence for different raking marginal variables (**VARDIFFTLR, DIFFTLR**) .

RAKING MACRO: GENERATION IV; PARAMATERS AND DEFAULTS

```
%MACRO RAKE_AND_TRIM_G4     /* new macroparameters are in red */
  (inds=,          /*    input data set             */
   outds=,         /*    output data set            */
   inwt=,          /*    input weight               */
   outwt=,         /*    raked weight               */
   varlist=,       /* list of raking variables      */
   numvar=,        /* number of raking variables    */
   trmprec=1,      /* termination criterion based on weighted total */
   cntotal = 100, /* if trmpct is used this is used to calculate weighted total*/
   trmpct=,        /* termination criterion based on marginal percent */
                   /* if present, the macro terminates based on this criterion */
   VARDIFFTLR =, /* list of variables with convergence criteria different from
                       trmpct; works only for trmpct criterion of convergence */
   DIFFTLR    =,  /* list of values of different tolerances */
   numiter=50,    /* maximum number of iterations */
   prdiag = N,    /* print detailed diagnostics? */
   NameRTF = ,    /* name of rtf output */

   TRIMWEIGHT = YES, /* switch: trim weight or not: YES or NO */
   METHOD = OR,      /* method of weight trim: AND or OR */
   GL_SWITCH = YES,        /* global low switch for OR method */
   GH_SWITCH = YES,        /* global high switch for OR method */
   IL_SWITCH = YES,        /* individual low switch for OR method */
   IH_SWITCH = YES,        /* individual high switch for OR method */
   GL_AND_IL_SWITCH = YES,   /* global and individual low switch for AND method */
   GH_AND_IH_SWITCH = YES,   /* global and individual high switch for AND method*/

   IHC      = 5,      /* weight will be decreased to individual weight times IHC*/
   ILC      = 0.2,    /* weight will be increased to individual weight times ILC */
   GHC      = 11.0,   /* weight will be decreased to mean weight times GHC  */
   GLC      = 0.091,  /* weight will be increased to mean weight times GLC  */
   INOC   = 15     /* iteration # to start checking on non-convergence*/)
```

You Do Not Have To Step On The Same Rake: SAS Raking Macro – Generation IV

David Izrael, Michael P. Battaglia, Ann A. Battaglia, Sarah W. Ball

TO TRIM OR NOT TO TRIM

The SAS raking macro has an “on/off switch” (**TRIMWEIGHT**) that allows the user to specify if weight trimming is to take place during the raking iterations. If no weight trimming is desired the user should specify “NO”.

BY WHICH METHOD TO TRIM

- The “OR” method (**METHOD=OR**) is the original trimming method developed. A case has its weight reduced if the value of the weight is greater than the individual high cap value or the global high cap value. A case has its weight increased if the value of the weight is less than the individual low cap value or the global low cap value.

- In the new “AND” method (**METHOD=AND**) a case has its weight reduced if the value of the weight is greater than the individual high cap value and the global high cap value. A case has its weight increased if the value of the weight is less than the individual low cap value and the global low cap value.

- The “OR” method will trim the weights of more cases than the “AND” method. It therefore allows for more control over extreme weights. However, if one has many raking margins and the weighted distributions prior to raking differ by a large degree from the control total distributions, convergence may be difficult to achieve. In that situation, the “AND” method may allow for reasonable control over extreme weight values while achieving convergence.

WHICH END TO TRIM

The user must next decide whether to trim weights on the “high end” and the “low end”, the “high end” only, or the “low end” only. This is done by setting the GL and IL switch to YES or NO, and setting the GH and IH switch to YES or NO. In most situations both switches should be set to YES. In some situations, the user may not be concerned about a small number of cases with very small weights especially if this is causing convergence problems. When that occurs the GL and IL switch should be set to NO.

BY HOW MUCH TO TRIM

- How does the user go about determining the trimming coefficients (**IHC, ILC, GHC, GLC** – in the previous versions they had names A, B, C, and D respectfully)? This is a two-step process. In the first step a UNIVARIATE on the raking input weight variable should be closely examined. The raking input weight distribution should be compared to the mean weight value to determine values for the global high cap value and the global low cap value.

- The choices for the individual high cap value and the individual low cap value are determined more by the preference of the user in terms of allowing individual weights to increase or decrease by too large an amount. Based on the initial specification of trimming values the raking macro is executed. From our experience the initial values (default) - **IHC=5, ILC=.2, GHC=11, GLC=.091** - work reasonably well.

- In the second step, if the raking fails to converge then one or more of the trimming values must be changed. If the raking converges one can still consider changing one or more of the trimming values to further reduce variability in the raked weights. One should review the raking macro output and examine a UNIVARIATE on the raked weight.

- If the first raking failed to converge then one or both high cap values needs to be increased, and one or both low cap values may need to be reduced. If the raking converged then one or both high cap values can be reduced, and one or both low cap values may be increased in order to further reduce extreme weights. After specifying new trimming values the raking macro is run for a second time and the results examined. This process may be repeated depending on the results of the second or subsequent rakings.

- The macro output includes the coefficient of variation of the raked weights. For comparison purposes one can run the raking macro with no weight trimming, and determine the degree to which the coefficient of variation of the raked weight has declined.

You Do Not Have To Step On The Same Rake: SAS Raking Macro – Generation IV

David Izrael, Michael P. Battaglia, Ann A. Battaglia, Sarah W. Ball

STUBBORN NON-CONVERGENCE USING PREVIOUS VERSION OF THE MACRO

```
%rake_and_trim
(inds= all,
outds=all_rake,
inwt= weight,
freqlist=,
outwt=Final_wgt,
varlist=  setting race gender age region occupation,
numvar=6,
cntotal=100,
trmprec=1000,
trmpct=0.1,
numiter=100,
prdiag=N,
namertf= raking,
MethTrim = IGCV,
A= 5,
B= 0.2,
C = 11.0,
D = 0.091,
INOC=25 );
```

The process never converged and never had a tendency to converge !

CONVERGENCE ACHIEVED WITH THE SAME DATA WITH NEW VERSION OF THE MACRO

```
%rake_and_trim_G4
(inds= all,
outds=all_rake,
inwt= weight,
outwt=Final_wgt,
varlist=  setting race gender age region occupation,
numvar=6,
cntotal=100,
trmprec=1000,
trmpct=0.5,
numiter=50,
prdiag=N,
namertf= raking,
TrimWeight = YES,
METHOD=AND,                  /* USING METHOD 'AND' */
GL_switch =,
GH_switch =,
IL_switch =,
IH_switch= ,
GL_and_IL_switch = YES,      /* USING BOTH LOW SWITCHES */
GH_and_IH_switch = YES,      /* USING BOTH HIGH SWTCHES */
IHC= 4,
ILC= 0.05,
GHC = 6.0,
GLC = 0.02,
INOC=25);
```

Convergence achieved in four iterations !

You Do Not Have To Step On The Same Rake: SAS Raking Macro – Generation IV

David Izrael, Michael P. Battaglia, Ann A. Battaglia, Sarah W. Ball

USING DIFFERENT CONVERGENCE CRITERIA FOR ONE OR MORE VARIABLES

```
%rake_and_trim_G4_V3
(inds= all,
outds=all_rake,
inwt= weight,
outwt=Final_wgt,
varlist=  setting race gender age region occupation,
numvar=6,
cntotal=100,
trmprec=1000,
trmpct=0.1,                          /* GENERAL CRITERION FOR ALL THE VARIABLES */
VARDIFFTLR = setting age occupation, /* WE WANT DIFFERENT CRITERIA FOR THREE VARIABLES*/
DIFFTLR   =   0.01   0.7   0.001,    /* FOR THE REST THREE 0.1 IS STILL A CRITERION */
numiter=50,
prdiag=N,
namertf= raking,
TrimWeight = YES,
METHOD=AND,
GL_switch =,
GH_switch =,
IL_switch =,
IH_switch= ,
GL_and_IL_switch = YES,
GH_and_IH_switch = YES,
IHC= 4,
ILC= 0.05,
GHC = 6.0,
GLC = 0.02,
INOC=25
);
```

WHAT ELSE IS NEW

- Header in the diagnostics includes new parameters:

  ……

  Trim weight?: **YES**
  Trimming method: **AND**
  GL and IL switch: **YES**
  General tolerance (percentage points): **0.1**
  Raking variable with different tolerance: **setting age occupation**
  Respective different tolerances: **0.01 0.7 0.001**

  ……

- Summary of trimmed cases by type of cap value:

  Number of Respondents Who Had Their Weights Decreased to Global High Cap Value (GHCV) : **0.**
  Number of Respondents Who Had Their Weights Increased to Global Low Cap Value (GLCV) : **0.**
  Number of Respondents Who Had Their Weights Decreased to Individual High Cap Value (IHCV) : **49**
  Number of Respondents Who Had Their Weights Increased to Individual Low Cap Value (ILCV) : **0.**

You Do Not Have To Step On The Same Rake: SAS Raking Macro – Generation IV

David Izrael, Michael P. Battaglia, Ann A. Battaglia, Sarah W. Ball

REFERENCES

Michael P. Battaglia, David Izrael, David C. Hoaglin, and Martin R. Frankel. Practical Considerations in Raking Survey Data. *Survey Practice, Online Journal, June 2009.* (Addendum with more information on weight trimming is forthcoming).

David Izrael, Michael P. Battaglia, Martin R. Frankel. Extreme Survey Weight Adjustment as a Component of Sample Balancing (a.k.a. Raking). *Proceedings of SAS Global Forum 2009,* March 2009, Washington, D.C.

Battaglia, M.P., Izrael, D., Hoaglin, D.C., and Frankel, M.R. 2004. Tips and tricks for raking survey data. *2004 Proceedings of the Annual Meeting of the American Statistical Association* [CD-ROM], Alexandria, VA: American Statistical Association.

David Izrael, David C. Hoaglin, and Michael P. Battaglia, “To Rake or Not To Rake Is Not the Question Anymore with the Enhanced Raking Macro “, *Proceedings of SUGI 2*9 2004, Montreal, Canada.

CONTACT

David Izrael, david_izrael@abtassoc.com, 617-349-2434

This methodology was developed and tested for CDC’s Surveys to Monitor Influenza Vaccination Coverage among Health Care Personnel and Pregnant Women during the 2014-15 Influenza Season project, Contract No. HHSD2002014M60311B, Order No.200-2014-F-61025

SAS® GLOBAL FORUM 2017

April 2 – 5 | Orlando, FL