



Array today

Utilizing the power of the array

Elsa Lozachmeur, Statistical programmer

Basel, 14th of October 2013

Objectives

Focus for today



WHY > HOW



Reminder about array

Data desc

Set

Array statment

Number of
elements

Array simple {4} r1c1-r1c4;

Do loop;

Array name

Array element names

SAS code (with rules to be applied);

End;

Run;

	COL1	COL2	COL3	COL4
ROW1	240	1000	72	95

Reminder about array

Data dsout;

Set dsin;

Dimension

**Array triple {2,3,4} l1r1c1-l1r1c4
l1r2c1-l1r2c4**

**Array double {3,4} r1c1-l1r3c4 r2c1-r2c4 r3c1-r3c4;
l2r1c1-l2r1c4
l2r2c1-l2r2c4
l2r3c1-l2r3c4;**

Do loop;

SAS code (with rules to be applied);

End;

Run;

LEVEL1:				
	COL1	COL2	COL3	COL4
ROW1	240	1000	72	95
ROW2	352	751	1204	56
ROW3	77	456	854	63
LEVEL2:				
	COL1	COL2	COL3	COL4
ROW1	785	435	946	7521
ROW2	356	857	125	67
ROW3	757	368	458	36

Background

■ Problem :

- complex rule to derive disease progression confirmation flag (repetitive calculations on a group of variables)
- Update, maintenance of code

■ Possible solutions:

- Macro (100+ lines of codes): [Slide 14](#) [Slide 15](#) [Slide 16](#)
- 2D-Array as an alternative solution

Clinical example

Derivation rule overview

Subject ID	Element	Array Reference	Values			Logic
02	Score	SC	3	3.5	3.5	a. Is there an increase in score compared to previous visit (Y), if yes then set onset to 1.
02	Onset	ON	1	2		b. Is the score sustained in the next visit (Y), if yes then set onset to 2.
02	Study Day	SD	-7	92	185	c. Get the difference in days between 2 visits when onset=2.
02	Difference	DIFF			93	
02	Visit	VIS	s	s	s	d. If the visit is a scheduled visit (s) and onset=2 (sustained) and difference greater than 90 then define disease progression as confirmed progression. The difference can be the sum of multiple visits.
02	Disease Progression Confirmed	DPCF			Y	

Clinical example

Details

```
data dsout; set dsintr;  
array dp{6,number of visit} SC SD VIS ON DIFF DPCF;
```

Subject ID	Element	Array Reference	Values			Logic
02	Visit	VIS	s	s	s	d. If the visit is a scheduled visit (s) and onset=2 (sustained) and difference greater than 90 then define disease progression as confirmed progression. The difference can be the sum of multiple visits.
02	Disease Progression Confirmed	DPCF			Y	

```
do i=1 to dim2(dp);
```

```
  if i gt 1 and i lt [number of visit] then do;
```

```
    if (dp{SC,i}) gt (dp{SC,i-1}) then dp{ON,i}=1;
```

```
    if (dp{SC,i+1}) ge (dp{SC,i}) and dp{ON,i} eq 1 then dp{ON,i+1}=2;
```

```
    if dp{ON,i}=2 then dp{DIFF,i}=sum(dp{SD,i}-dp{SD,i-1},dp{DIFF,i-1});
```

```
    if dp{VIS,i}="s" and dp{DIFF,i} gt 90 then dp{DPCF,i}="Y";
```

```
  end;
```

```
end;
```

```
run;
```

Clinical example

SAS Code

```
data dsout;
  set dsintr;

  array dp{6,15} SC1-SC15 SD1-SD15 VIS1-VIS15 ON1-ON15 DIFF1-DIFF15 DPCF1-DPCF15;

  do i=1 to dim2(dp);

    if i gt 1 and i lt 15 then do;

      if dp{1,i}) gt (dp{1,count-1}) then dp{4,i}=1;
      if (dp{1,count+1}) ge (dp{1,i}) and dp{4,i} eq 1 then dp{4,i+1}=2;
      if dp{4,i}=2 then p{5,i}=sum(dp{2,i}-dp{2,i-1},dp{5,i-1});
      if dp{3,i}="s" and dp{5,i} gt 90 then dp{6,i}="Y";

    end;

  end;

run;
```

Benefits

- After the submission, we were asked to implement new, stricter rules in our codes, by adding new parameters.

```
data dsout;  
  set dsintr;  
  Increment number of elements 5 vis1-vis15 on1-on15 diff1-diff15 dpcf1-dpcf15 vnb1-vnb15  
  Get the day of the onset visit 15 timeon1-timeon15 timecf1-timecf15;  
  do i=1 to dim2(dp);  
    if i gt 1 and i lt 15 then do;  
      if (dp{1,i}) at (dp{1,i-1}) then dp{4,i}=1;  
      Get visit 15 if dp{4,i}=1 then dp{7,i}=dp{3,i};  
      if (dp{1,i+1}) ge (dp{1,i}) and dp{4,i} eq 1 then dp{4,i+1}=2;  
      if dp{4,i}=2 then dp{5,i}=sum(dp{2,i}-dp{2,i-1},dp{5,i-1});  
      if dp{3,i}="s" and dp{5,i} gt 90 then dp{6,i}=1;  
      New and stricter rule 15 if dp{5,i} ge 76 then dp{8,i}=1;  
      Get the day of the onset visit 2,i);  
      Get the day of the confirmation visit 2,i);  
    end;  
  end;  
  Get the total number of confirmation  
run;
```

Benefits

Summary

■ Technically:

- Reduce lines of code
- Easier to maintain
- Easier to add new criteria

■ Personally:

- Don't fear arrays
- New tool
- Encourage other people to use it
- Increase network within the company

What's next?

- Improvement:
 - use arrays to transform data
 - use option `_TEMPORARY_` to not provide name for elements
- A sublevel is added (e.g. Right/left side for the score): don't worry the third dimension exists!
- Increase the appetite to learn more about do loop and arrays.
- Be dynamic:
 - use `dim` function
 - use macro variable to define the dimension

Conclusion



Acknowledgments

- Thanks to:
- Co-author: Joseph Rowley
- SAS code improvement help: Michel Gauthier
- Reviewers: Anne-Lise Illy, Rohit Kamath, James Gallagher, Wang Zhang

Macro code

```
%macro disprog(select=,dayconf=,dout=,sortsustain=,select2=%str(conf=1),cut=cutm24,ty=1);
** select = Selection criteria for the record to be assessed for disease progression;
** daycrit = Number of days before disease progression can be confirmed;
** dout = output dataset;
** sortsustain = sort order for determining last sustained visit e.g descending cut24;
** select2 = default is to report confirmed status.;
** cut= cut used for determining if dp is sustained;
** ty= 1= assess for progression , 2 = assess for improvement;

data dpat; ** Patients who could have disease progression or at least have a record that meets onset criteria;
set visdates;

%if &ty.=1 %then %do;
** flag records that meet DP onset and progression (not confirmed);

    if 0<=exdsc1nb<=5 and EXDNCHG >=1 then DPNC=1;
    if exdsc1nb>= 5.5 and EXDNCHG >=0.5 then DPNC=1;
    if EXDSC1N = 10 then DPNC=1;** Patient dies due to MS;
%end;
%if &ty.=2 %then %do;
** flag records that meet NI onset and improvement (not confirmed);

    if 1<=exdsc1nb<=5.5 and not missing(EXDNCHG) and EXDNCHG <=-1 then DPNC=1;
    if exdsc1nb>=6 and not missing(EXDNCHG) and EXDNCHG <=-0.5 then DPNC=1;
%end;

run;

proc sort data = dpat out = dpats(keep =stysid1a) nodupkey;
by stysid1a ;
where DPNC=1;
run;

data dp(keep=stysid1a exmdat1o cutoff2o cut12M2O vis1n rpevis1n exdsc1nb exdsc1n EXDNCHG dpnc vis2n dayni_1n fseprf1c where=(not missing(exdsc1n)));
merge dpat(in=a %select.) dpats(in=b);
by stysid1a;
if b;
run;

proc sort data=dp out =dp1;
by stysid1a exmdat1o vis1n rpevis1n vis2n;
run;
```

Macro code

```
data dp2;
set dp1;

by stysid1a exmdat1o vis1n rpevis1n vis2n;
daylag=lag(dayni_1n);
dplag=lag(dpnc);
if first.stysid1a then do;
    daylag=.;dplag=.;
end;
if dpnc=1 and dplag=. then daylag=.;
if not missing(daylag) and dpnc =1 then daydif=dayni_1n - daylag;
** DNCP flag records that meet onset and progression (not confirmed);
** DAYLAG Value DAYNI_1N for previous record that meet DPNC =1;
** DPLAG Does the record have a previous record no gap meeting DPNC?;
** DAYDIF Number of days has the record met DPNC;
** DAYPROG CUMULATIVE days patients has had DPNC no gap;
** CONF Does the record meet confirmation criteria?;

run;

data dp3;
set dp2;
retain dayprog sus onset oscore;
by stysid1a;

** identify vis2n scheduled visits;
**sched=mod(vis2n,1); ** 0 = scheduled;

if first.stysid1a then dayprog=0;
else dayprog+daydif;

if daydif=. then dayprog=0;
if FSEPRF1C ne 1 and dayprog >= &dayconf. and vis1n ne 999 then conf=1;

if nmiss(exmdat1o,cutoff2o) = 0 then cutm24=datepart(exmdat1o)<=datepart(cutoff2o);
if nmiss(exmdat1o,cut12M2O) = 0 then cutm12=datepart(exmdat1o)<=datepart(cut12M2O);

** flag if all visits after confirmed meet dp criteria;
if first.stysid1a then sus=.;
if conf=1 then sus =1;
else if dpnc ne 1 then sus=0;
else if cutm24 = 0 then sus=0;
```

Macro code

```
** Get onset and score date;
if first.stysid1a then onset=.;
if dpnc=1 and daylag=. then do; onset=dayni_1n; oscore=EXDSC1N; end;
else if dpnc ne 1 then do; onset=.; oscore=.; end;
run;

*** sustained to last visit?;
proc sort data = dp3 out = sustain1 ;
  by stysid1a $sortsustain dayni_1n;
run;

data sustain2;
set sustain1;
by stysid1a $sortsustain. ;
if $select2 ne $str(conf=1) %then %do;
  if last.&cut. and sus=1 then sustain=1;
%end;

*** if patient has predose value set onset to 1;
if not missing(onset) and onset <1 then onset =1;
*** Keep date patient had confirmed visit;
*** if patient has predose value set onset to 1;
if not missing(dayni_1n) and $select2. then confday = dayni_1n;
cscore = EXDSC1N; ** confirmation score;
run;

proc sort data = sustain2 out = $dout. (keep=stysid1a onset confday cscore oscore) nodupkey;
  by stysid1a;
  where $select2. ;
run;

%mend disprog;

%disprog(select=$str(where =(flagfse in (1,2) and dayni_1n >=1)),dayconf=76,dout=cdp3mfas); ** 3 m confirmed dp (FAS);
%disprog(select=$str(where =(flagfse in (1,2) and dayni_1n >=1)),dayconf=166,dout=cdp6mfas); ** 6 m confirmed dp (FAS);
%disprog(select=$str(where =(flagfse in (1) and dayni_1n >=1)),dayconf=76,dout=cdp3mfp); ** 3 m confirmed dp (PP);
%disprog(select=$str(where =(flagfse in (1,2) and dayni_1n >=1)),dayconf=76,dout=cdp3msus,sortsustain=$str(descending cutm24),select2=$str(sustain=1)); ** 3 m confirmed dp (FAS);
%disprog(select=$str(where =(flagfse in (1,2) and dayni_1n >=1)),dayconf=166,dout=cdp6msus,sortsustain=$str(descending cutm24),select2=$str(sustain=1)); ** 6 m confirmed dp (FAS);
%disprog(select=$str(where =(flagfse in (1,2) and dayni_1n >=1)),dayconf=76,dout=cni3mfas,ty=2); ** 3 m confirmed NI (FAS);
%disprog(select=$str(where =(flagfse in (1,2) and dayni_1n >=1)),dayconf=166,dout=cni6mfas,ty=2); ** 6 m confirmed NI (FAS);
```

