PharmaSUG 2017 - Paper QT11

Multi-Dimensional Arrays: Add Derived Parameters

Siddharth Kumar Mogra, GCE Solutions Inc

ABSTRACT

Multidimensional arrays are useful when Data needs to be grouped into a 'table like' arrangement for processing. Often we are required to derive and add records in analysis datasets. When there are large numbers of parameters to be added, the use of repeated if-then statements are prone to error and time consuming. Effective use of Multi-Dimensional Arrays or Nested Arrays can increase efficiency of program. It can minimize the code and makes program efficient. This paper demonstrates situation where use of Multi-Dimensional Arrays is practical and appropriate in Clinical Trials.

INTRODUCTION

Creating Laboratory analysis datasets is challenging and daunting task, due to the size of dataset and number of derived parameters required. Creating Lab analysis dataset requires to derive parameters. Common technique used to add records in dataset is the use of if-then statement, or create dummy dataset etc. Use of Nested arrays to derive these parameters limits the chance of error and makes program efficient.

Arrays can be very powerful with its iterative and conditional processing. Array in SAS® allows to group a bunch of variables for the same process. The huge block of repetitious statements and redundant calculation codes can be reduced to just a few lines. With arrays you can simplify the coding in many cases and can accomplish tasks which are not easily done.

An array must be defined within the data step prior to being referenced. if an array is referenced without first being defined, an error will occur. As array exist only for the duration of the data step in which they are defined therefore, they can only be used in the same data step where it is referenced.

WHY ARRAYS

Arrays are useful and effective tool when we want to perform same actions on multiple variables. Arrays allows to group a bunch of variables for the same process. The huge block of the repetitious statements and redundant calculation codes can be reduced to just a few lines. Code can be simplified. Some of the examples where use of arrays in clinical trials are, to derive / Update Treatment Variables in Analysis Dataset, to derive / Update Analysis Flag Variables (anl01fl,) in Analysis Dataset , to concatenate and Apply Format to all the Treatment Variables for presentation in Output, etc.

THE SYNTAX:

array array-name {n} <$> <length> array-elements <(initial values)>;

- array-name – Any valid SAS name that identifies the group of variables
- n – Number of elements within the array
- $ - Indicates the elements within the array are character type variables
- Length – assigns length for the array elements
- Elements – List of SAS variables to be part of the array
- Initial values – Provides the initial values for each of the array elements.

ARRAY REFERENCE:

array-name {*subscript*}

- Where *array-name* - is the name of an array that was previously defined with ARRAY statement in the same DATA step.
- *Subscript* - specifies the subscript, which can be a numeric constant, the name of a variable whose value is the number, a SAS numeric expression, or an asterisk (*).
- An array must be defined within the data step prior to being referenced.
- Array exists only for the duration of the data step in which they are defined

ONE DIMENSIONAL ARRAYS

A simple array may be created when the variables grouped together conceptually appear as a single row. This is known as a one-dimensional array.

The array statement to define one-dimensional array will be:

```
array temperature_array {24} temp1 – temp24;
```

The array has 24 elements for the variables TEMP1 through TEMP24. When the array elements are used within the data step the array name and the element number will reference them.
The reference to the ninth element in the temperature array is: temperature_array{9}
The statement used to define an array is the ARRAY statement.

```
array array-name {n}  <$>  <length> array-elements  <(initial-values)>;
```

The Array statements provides the following information about SAS array:

array-name – Any valid SAS name

n – Number of elements within the array

$ - Indicates the elements within the array are character type variables

Length – assigns length for the array elements

Elements – List of SAS variables to be part of the array

Initial values – Provides the initial values for each of the array elements.

In clinical trials arrays are mostly applied in the creation of TLFs. Example of situation where arrays are commonly used are: Creating final format on treatment columns

MULTI DIMENSIONAL ARRAY

Multi dimensional arrays are used when you want to put values in a table format (i.e., rows and columns). Multidimensional arrays can be used to input data or to perform operations on the data set.

Multi dimensional arrays are defined as follows: The number of elements are placed in each dimension after the array name in the form {n, ..}. From right to left, the rightmost dimension represents columns; the next dimension represents rows. Each position farther left represents a higher dimension.

```
array temprg {3, 5} c1t1 c1t2 c1t3 c1t4 c1t5
                    c2t1 c2t2 c2t3 c2t4 c2t5
```

c3t1 c3t2 c3t3 c3t4 c3t5;

This is an example of two-dimensional array with three rows and five columns The array contains fifteen variables: five temperature measures (t1 through t5) from three cities (c1, c2 and c3).

SAS ® places variables into a multidimensional array by filling all rows in order, beginning at the upper left corner of the array (known as row-major order). You can think of the variables as having the following arrangement:

c1t1 c1t2 c1t3 c1t4 c1t5
c2t1 c2t2 c2t3 c2t4 c2t5

To refer to an element of the array later with an array reference, you can use the array name and subscripts. Always specify the row you want first, and then the column. Separate the two numbers with a comma. The following table lists some of the array references for the previous example:

Array References for array TEMPRG	
Variable	Array Reference
c1t1	temprg{1,1}
c1t2	temprg{1,2}
c2t2	temprg{2,2}
c2t5	temprg{2,5}

In summary, listing the dimensions of arrays works like the following:

One-dimensional array: array x(cols)

Two-dimensional array: array y(rows,cols)

Three-dimensional array: array z(levels, rows, cols)

USING NESTED DO LOOPS

Multidimensional arrays are usually processed inside nested Do loops.

For example, a two-dimensional array can be processed as follows:

```
Do index-variable-1 = 1 to number of rows;
    Do index-variable-2=1 to number of columns;
        more SAS statements …
    End;
End;
```

An array reference can use two or more index variables as the subscript to refer to two or more

dimensions of an array. Use the following form:

array-name {*index-variable-1, …,index-variable-n*}

The following example creates an array that contains fifteen variables- five temperature measures (t1 through t5) from three cities (c1, c2, c3). The DATA step contains two DO loops.

A do loop is needed for each dimension, thus two do loops are specified below, one for the rows (which is represented by i and set from 1 to 3) and one for the columns (represented by j and set from 1 to 5). Note, if you make i reference the rows (1 to 3), then i should be put in the first position in the array.

The outer DO loop (DO I=1 TO 3) processes the inner DO loop thrice. The inner DO loop (DO J=1 TO 5) applies the ROUND function to all the variables in one row. For each iteration of the DO loops, SAS substitutes the value of the array element corresponding to the current values of I and J.

```
options linesize=80 pagesize=60;

data temps;
   array temprg{3,5} c1t1-c1t5 c2t1-c2t5 c3t1-c3t5;
   do i=1 to 3;
     do j=1 to 5;
       temprg{i,j}=round(temprg{i,j});
     end;
   end;
   run;
```

The previous example can also use the DIM function to produce the same result:

```
do
i=1 to dim1(temprg);
   do j=1 to dim2(temprg);
      temprg{i,j}=round(temprg{i,j});
   end;
end;
```

The value of DIM1(TEMPRG) is 3; the value of DIM2(TEMPRG) is 5.

APPLICATION IN CLINICAL TRIAL

An example of application of Two-Dimensional arrays to derive parameters in laboratory analysis dataset

PROGRAMMING SPECIFICATIONS

As per the Programming specifications, 24 parameters are required to add in ADLB (lab) dataset. Table 1, is an example of the parameters required.

Table 1: Programming Specifications

PARCAT1	PARAMCD	PARAM	PARAMN	PARCAT4	PARAMTYP	DTYPE
CHEMISTRY	ALP2U	ALP > 2 x ULN		ALPSI	DERIVED	COPY
CHEMISTRY	ALP3U	ALP > 3 x ULN		ALPSI	DERIVED	COPY
CHEMISTRY	ALP5U	ALP > 5 x ULN		ALPSI	DERIVED	COPY
CHEMISTRY	ALT10U	ALT > 10 x ULN		ALTSI	DERIVED	COPY
CHEMISTRY	ALT20U	ALT > 20 x ULN		ALTSI	DERIVED	COPY
CHEMISTRY	ALT3U	ALT > 3 x ULN		ALTSI	DERIVED	COPY
CHEMISTRY	ALT5U	ALT > 5 x ULN		ALTSI	DERIVED	COPY
CHEMISTRY	ALT8U	ALT > 8 x ULN		ALTSI	DERIVED	COPY
CHEMISTRY	AST10U	AST > 10 x ULN		ASTSI	DERIVED	COPY
CHEMISTRY	AST20U	AST > 20 x ULN		ASTSI	DERIVED	COPY
CHEMISTRY	AST3U	AST > 3 x ULN		ASTSI	DERIVED	COPY
CHEMISTRY	AST5U	AST > 5 x ULN		ASTSI	DERIVED	COPY
CHEMISTRY	AST8U	AST > 8 x ULN		ASTSI	DERIVED	COPY
CHEMISTRY	TBL1_5U	TBL > 1.5 x ULN		BILISI	DERIVED	COPY
CHEMISTRY	TBL2U	TBL > 2 x ULN		BILISI	DERIVED	COPY
CHEMISTRY	TBL3U	TBL > 3 x ULN		BILISI	DERIVED	COPY

DATA:

Dataset 1: Lab Analysis Dataset

	SUBJID	AVISITN	PARAM	PARAMCD	PARAMTYP	PARCAT1	AVAL	AVALC	TRTA	R2ANRHI	R2ANRLO	A
1	4001299	4	Albumin (g/L)	ALBSI		CHEMISTRY	45	45	Treatment	0.8653846154	1.2857142857	V
2	4001299	4	Alkaline phosphatase (U/L)	ALPSI		CHEMISTRY	60	60	Treatment	0.4651162791	1.5	V
3	4001299	4	Alanine aminotransferase (U/L)	ALTSI		CHEMISTRY	21	21	Treatment	0.512195122	.	V
4	4001299	4	Aspartate aminotransferase (U/L)	ASTSI		CHEMISTRY	18	18	Treatment	0.4864864865	.	V
5	4001299	4	Bilirubin (umol/L)	BILISI		CHEMISTRY	9	9	Treatment	0.4166666667	2.5	V
6	4001299	4	Creatinine (umol/L)	CREATSI		CHEMISTRY	97	97	Treatment	0.9166666667	1.5714285714	V
7	4001299	4	Phosphate (mmol/L)	PHOSSI		CHEMISTRY	1.07	1.07	Treatment	0.7333333333	1.2222222222	V
8	4001299	8	Albumin (g/L)	ALBSI		CHEMISTRY	41	41	Treatment	0.7884615385	1.1714285714	V
9	4001299	8	Alkaline phosphatase (U/L)	ALPSI		CHEMISTRY	66	66	Treatment	0.511627907	1.65	V
10	4001299	8	Alanine aminotransferase (U/L)	ALTSI		CHEMISTRY	18	18	Treatment	0.4390243902	.	V
11	4001299	8	Aspartate aminotransferase (U/L)	ASTSI		CHEMISTRY	21	21	Treatment	0.5675675676	.	V
12	4001299	8	Bilirubin (umol/L)	BILISI		CHEMISTRY	9	9	Treatment	0.4166666667	2.5	V
13	4001299	8	Creatinine (umol/L)	CREATSI		CHEMISTRY	120	120	Treatment	1.1333333333	1.9428571429	V
14	4001299	8	Phosphate (mmol/L)	PHOSSI		CHEMISTRY	1	1	Treatment	0.6888888889	1.1481481481	V
15	4001299	12	Albumin (g/L)	ALBSI		CHEMISTRY	41	41	Treatment	0.7884615385	1.1714285714	V
16	4001299	12	Alkaline phosphatase (U/L)	ALPSI		CHEMISTRY	61	61	Treatment	0.4728682171	1.525	V
17	4001299	12	Alanine aminotransferase (U/L)	ALTSI		CHEMISTRY	15	15	Treatment	0.3658536585	.	V
18	4001299	12	Aspartate aminotransferase (U/L)	ASTSI		CHEMISTRY	17	17	Treatment	0.4594594595	.	V
19	4001299	12	Bilirubin (umol/L)	BILISI		CHEMISTRY	9	9	Treatment	0.4166666667	2.5	V
20	4001299	12	Creatinine (umol/L)	CREATSI		CHEMISTRY	107	107	Treatment	1.0083333333	1.7285714286	V
21	4001299	12	Phosphate (mmol/L)	PHOSSI		CHEMISTRY	1.19	1.19	Treatment	0.8222222222	1.3703703704	V
22	4001299	16	Albumin (g/L)	ALBSI		CHEMISTRY	44	44	Treatment	0.8461538462	1.2571428571	V
23	4001299	16	Alkaline phosphatase (U/L)	ALPSI		CHEMISTRY	65	65	Treatment	0.503875969	1.625	V
24	4001299	16	Alanine aminotransferase (U/L)	ALTSI		CHEMISTRY	16	16	Treatment	0.3902439024	.	V
25	4001299	16	Aspartate aminotransferase (U/L)	ASTSI		CHEMISTRY	17	17	Treatment	0.4594594595	.	V
26	4001299	16	Bilirubin (umol/L)	BILISI		CHEMISTRY	5	5	Treatment	0.25	1.5	V
27	4001299	16	Creatinine (umol/L)	CREATSI		CHEMISTRY	99	99	Treatment	0.9333333333	1.6	V
28	4001299	16	Phosphate (mmol/L)	PHOSSI		CHEMISTRY	0.94	0.94	Treatment	0.6444444444	1.0740740741	V

USE OF NESTED ARRAYS

Here is an example of use of nested arrays where the parameters that are required as per the programming specifications are derived.

Output 1: Ex. Multi-Dimensional Array

```
data dlb2;
    set dlb1;
    array _param(4,5) $200. ('ALT > 3 x ULN'    'ALT > 5 x ULN' 'ALT > 8 x ULN' 'ALT > 10 x ULN' 'ALT > 20 x ULN'
                             'AST > 3 x ULN'    'AST > 5 x ULN' 'AST > 8 x ULN' 'AST > 10 x ULN' 'AST > 20 x ULN'
                             'TBL > 1.5 x ULN'  'TBL > 2 x ULN' 'TBL > 3 x ULN' ''                ''
                             'ALP > 2 x ULN'    'ALP > 3 x ULN' 'ALP > 5 x ULN' ''                '');
    array _paramcd(4,5) $8. ('ALT3U'   'ALT5U' 'ALT8U' 'ALT10U' 'ALT20U'
                             'AST3U'   'AST5U' 'AST8U' 'AST10U' 'AST20U'
                             'TBL1_5U' 'TBL2U' 'TBL3U' ''       ''
                             'ALP2U'   'ALP3U' 'ALP5U' ''       '');
    array _ratio_limit(4,5) (3    5 8 10 20
                             3    5 8 10 20
                             1.5  2 3  .  .
                             2    3 5  .  .);
    array _parcat4 (4) $20. ('ALTSI' 'ASTSI' 'BILISI' 'ALPSI');
    output;
      AVAL=.;
      PARAMTYP='DERIVED'; DTYPE='COPY';

        do j=1 to 4;
          if paramcd=_parcat4(j) then do;

                do i=1 to 5;
                    if _param(j,i) ne '' then do;
                    param=_param(j,i);
                    paramcd=_paramcd(j,i);
                    if R2ANRHI>_ratio_limit(j,i)>. then do; avalc = 'Y'; aval = 1; end;
                    else if .<R2ANRHI<=_ratio_limit(j,i) then do; avalc = 'N'; aval = 0; end;
                    output;
                end;
            end;
          end;
        end;
run;
```

Output 2: Dataset with Added Parameters

SUBJID	AVISITN	PARAM	PARAMCD	PARAMTYP	PARCAT1	AVAL	AVALC	R2ANRHI	R2ANRLO	TRTA	DTYPE
4001299	4	Albumin (g/L)	ALBSI		CHEMISTRY	45	45	0.8653846154	1.2857142857	Treatment	
4001299	4	Alkaline phosphatase (U/L)	ALPSI		CHEMISTRY	60	60	0.4651162791	1.5	Treatment	
4001299	4	ALP > 2 x ULN	ALP2U	DERIVED	CHEMISTRY	0	N	0.4651162791	1.5	Treatment	COPY
4001299	4	ALP > 3 x ULN	ALP3U	DERIVED	CHEMISTRY	0	N	0.4651162791	1.5	Treatment	COPY
4001299	4	ALP > 5 x ULN	ALP5U	DERIVED	CHEMISTRY	0	N	0.4651162791	1.5	Treatment	COPY
4001299	4	Alanine aminotransferase (U/L)	ALTSI		CHEMISTRY	21	21	0.512195122	.	Treatment	
4001299	4	ALT > 3 x ULN	ALT3U	DERIVED	CHEMISTRY	0	N	0.512195122	.	Treatment	COPY
4001299	4	ALT > 5 x ULN	ALT5U	DERIVED	CHEMISTRY	0	N	0.512195122	.	Treatment	COPY
4001299	4	ALT > 8 x ULN	ALT8U	DERIVED	CHEMISTRY	0	N	0.512195122	.	Treatment	COPY
4001299	4	ALT > 10 x ULN	ALT10U	DERIVED	CHEMISTRY	0	N	0.512195122	.	Treatment	COPY
4001299	4	ALT > 20 x ULN	ALT20U	DERIVED	CHEMISTRY	0	N	0.512195122	.	Treatment	COPY
4001299	4	Aspartate aminotransferase (U/L)	ASTSI		CHEMISTRY	18	18	0.4864864865	.	Treatment	
4001299	4	AST > 3 x ULN	AST3U	DERIVED	CHEMISTRY	0	N	0.4864864865	.	Treatment	COPY
4001299	4	AST > 5 x ULN	AST5U	DERIVED	CHEMISTRY	0	N	0.4864864865	.	Treatment	COPY
4001299	4	AST > 8 x ULN	AST8U	DERIVED	CHEMISTRY	0	N	0.4864864865	.	Treatment	COPY
4001299	4	AST > 10 x ULN	AST10U	DERIVED	CHEMISTRY	0	N	0.4864864865	.	Treatment	COPY
4001299	4	AST > 20 x ULN	AST20U	DERIVED	CHEMISTRY	0	N	0.4864864865	.	Treatment	COPY
4001299	4	Bilirubin (umol/L)	BILISI		CHEMISTRY	9	9	0.4166666667	2.5	Treatment	
4001299	4	TBL > 1.5 x ULN	TBL1_5U	DERIVED	CHEMISTRY	0	N	0.4166666667	2.5	Treatment	COPY
4001299	4	TBL > 2 x ULN	TBL2U	DERIVED	CHEMISTRY	0	N	0.4166666667	2.5	Treatment	COPY
4001299	4	TBL > 3 x ULN	TBL3U	DERIVED	CHEMISTRY	0	N	0.4166666667	2.5	Treatment	COPY
4001299	4	Creatinine (umol/L)	CREATSI		CHEMISTRY	97	97	0.9166666667	1.5714285714	Treatment	
4001299	4	Phosphate (mmol/L)	PHOSSI		CHEMISTRY	1.07	1.07	0.7333333333	1.2222222222	Treatment	
4001299	8	Albumin (g/L)	ALBSI		CHEMISTRY	41	41	0.7884615385	1.1714285714	Treatment	
4001299	8	Alkaline phosphatase (U/L)	ALPSI		CHEMISTRY	66	66	0.511627907	1.65	Treatment	
4001299	8	ALP > 2 x ULN	ALP2U	DERIVED	CHEMISTRY	0	N	0.511627907	1.65	Treatment	COPY
4001299	8	ALP > 3 x ULN	ALP3U	DERIVED	CHEMISTRY	0	N	0.511627907	1.65	Treatment	COPY
4001299	8	ALP > 5 x ULN	ALP5U	DERIVED	CHEMISTRY	0	N	0.511627907	1.65	Treatment	COPY

BENEFITS:

Use of Multidimensional arrays reduces lines of code, and is easier to maintain. It is less prone to error, the review of code is quicker and the program is efficient.

CONCLUSION

When there are large number of parameters to derive, SAS two-dimensional arrays can provide a great deal of flexibility and increases the efficiency of program.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Siddharth Kumar Mogra
GCE Solutions, Inc
DLF Towers, New Delhi
New Delhi 110001
Work Phone: 9727583226
Email: Siddharth.kumar@gcesolutions.com