

Arrays in the Data Step – When and Why?

Ralf Minkenber

ARRAY *array-name*{*subscript*} <\$> <*length*> <*array-elements*> <(initial-value-list)> ;

array-name

Name of the Array

- identifies the Array
- user-defined name
- no SAS function name
- distinguishable from all variable names
- isn't a variable

ARRAY

Command for initialization of an Array

- defines a variable set as Array elements
- simple method to identify a group of variables temporarily (numeric / character)
- (only) exists inside a DATA step
- isn't a data structure

ARRAY *array-name*{*subscript*} <\$> <*length*> <*array-elements*> <(initial-value-list)> ;

{subscript}

number and order of the Array elements

- defines the number of elements in the Array
- describes the elements' order
- determines the dimension of the Array
- can contain a number, many numbers or a *

{ Dimension } array simple{3}; array x{5,3};

{ <lower:>upper } array x{1:5,1:3};

{*}

- embedded in brackets {} [] ()

array-elements

variable names

- must be all numeric or all character
- in addition: `_all_` `_numeric_` `_character_`
`_temporary_`

(initial-value-list)

initial values for the Array elements

Array definitions

```
array rain {5} janr febr marr aprr mayr ;  
array month {*} jan feb jul okt nov ;  
array meal {3} ;  
array score{5,3} score1-score15 ;  
array test {3:4,3:7} test1-test10 ;
```

<i>Array</i>	<i>Variable</i>	<i>Value</i>
rain {3}	marr	
month {5}	nov	
meal {1}	meal1	
score {3,2}	score8	
test {4,4}	test7	

Array definitions with initial values

```
array test {4} t1-t4 (90 80 2*70) ;  
array _time {5} _temporary_ (0 3 7 14 35) ;  
array x {10} a b c d e f g h i j (2*1:5) ;
```

test {4}	t4	70
_time {3}		7
x {7}	g	2

```
data bsp1;  
  array days {7} ;  array hours {7} h1-h7 ;  
  do i=1 to 7 ;  
    if days {i} = 99 then days {i} = 100 ;  
    hours {i} = days {i} * 24 ;  
  end ;  
run;
```

```
data bsp2;  
  array cost {5} c1 f1 k1 k2 z1 ;  
  totcost = sum (of cost {*}) ;  
  
  array big {3,4} wt1-wt4 ht1-ht4 bmi1-bmi4 ;  
  do i=1 to dim(big,2) ;  
    big {3,i} = big{1,i} / big{2,i}**2 ;  
  end ;  
run;
```

Explicit:

```
array item {1:6} xvar yvar zvar xxvar yyvar zzvar ;  
do i=1 to 6 ;    item {i} = item {i} / 100 ;    end ;
```

Implicit:

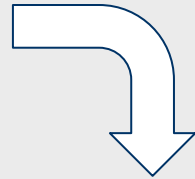
```
array item xvar yvar zvar xxvar yyvar zzvar ;  
do over item ;    item = item / 100 ;    end ;
```

Implicit and explicit:

```
array item xvar yvar zvar ;  
array results(1:3) ;  
do over item ; if item > 3.5 then results {_i_} = 1 ; end ;
```

An Example with an Implicit Array

ID	VAL1	VAL2	VAL3	VAL4	VAL5	VAL6
1	4	-1	3	3	.	.
2	-1	2	6	5	-1	2
3	0	1	3	2	5	.

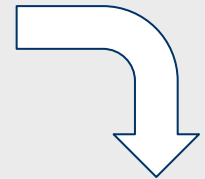


```
data test ;  
  set original ;  
  array values val: ;  
  n_of_v = 0 ; last_of_v = 0 ;  
  do over values ;  
    if values < 0 then values = . ;  
    else n_of_v + 1 ;  
    if values ^= . then last_of_v = _i_ ;  
  end ;  
run ;
```

ID	N_OF_V	LAST_OF_V
1	3	4
2	4	6
3	5	5

Changing the Data Structure Using Arrays

PATNO	HTCM1	HTCM2	HTCM3	WTKG1	WTKG2	WTKG3
10001	122	124	124	43	44.5	45
10002	109	112	117	33.5	36	39
10003	115	115	119	40	39	46



```
data vertical ;  
  set horizontal ;  
  array ahtcm {1:3} htcm1-htcm3 ;  
  array awtkg {1:3} wtkg1-wtkg3 ;  
  do timept=1 to 3 ;  
    bmi = awtkg {timept} /  
          (ahtcm {timept} / 100) ** 2 ;  
    htcm = ahtcm {timept} ;  
    wtkg = awtkg {timept} ;  
    output ;  
  end ;  
  keep timept patno htcm wtkg bmi ;  
  format bmi 4.1 ;  
run ;
```

PATNO	TIMEPT	HTCM	WTKG	BMI
10001	1	122	43	28.9
10001	2	124	44.5	28.9
10001	3	124	45	29.3
10002	1	109	33.5	28.2
10002	2	112	36	28.7
10002	3	117	39	28.5
10003	1	115	40	30.2

Error Messages Related to Arrays

- ERROR: Undeclared array referenced: ____.
- ERROR: Variable ____ has not been declared as an array.
- ERROR: Array subscript out of range at line ____ column ____.
- ERROR: Attempt to initialize variable ____ in numeric array ____ with character constant.
- ERROR: Too few array subscripts specified for array ____.
- ERROR: Too many array subscripts specified for array ____.
- ERROR: Illegal reference to the array ____.
- ERROR: Too few variables defined for the dimension(s) specified for the array ____.
- ERROR: The array ____ has been defined with zero elements.
- WARNING: Partial value initialization of the array ____.
- WARNING: Too many values for initialization of the array ____.
Excess values are ignored.
- NOTE: The array ____ has the same name as a SAS-supplied or user-defined function. Parentheses following this name are treated as array references and not function references.

When Arrays are Helpful?

- Repetitive, interrelated variables
- Many identical statements for a large set of variables
- To image complex relations between variables
- Re-structuring of (huge) amount of variables
- Use of a huge number of constant values

Arrays aren't a mystery and can produce more efficient and readable programs.