

2 PROC TRANSPOSES =
1 DATA Step DOW-Loop

Michigan SAS Users Group Meeting

3/11/2004

Nancy Brucken

STATPROBE, Inc.

Problem

- Original dataset (sorted by ANALYTEC and LISTNUM):

ANALYTEC	LISTTYPE	LISTNUM	RXGRP	PATCT	DENOM
CALCIUM	Subjects With One or More Abnormal Screen/Baseline Values	1	1	6	260
CALCIUM	Subjects With One or More Abnormal Screen/Baseline Values	1	2	4	227
CALCIUM	Subjects With Normal Screen/Baseline Values and One Abnormal Value After Initiation of Treatment	2	1	5	260
CALCIUM	Subjects With Normal Screen/Baseline Values and One Abnormal Value After Initiation of Treatment	2	2	6	227

Goal

- Desired dataset:

ANALYTEC	LISTTYPE	LISTNUM	PATCT1	PATCT2	DENOM1	DENOM2
CALCIUM	Subjects With One or More Abnormal Screen/Baseline Values	1	6	4	260	227
CALCIUM	Subjects With Normal Screen/Baseline Values and One Abnormal Value After Initiation of Treatment	2	5	6	260	227

PROC TRANSPOSE Limitation

- Multiple variables in VAR statement → one record per variable in output dataset

Solution

- Run a separate PROC TRANSPOSE for each variable, and then merge resulting datasets.

Drawbacks

- Requires a separate pass through the dataset for each PROC TRANSPOSE
- Requires another pass through each output dataset for the merge

Alternate Solution

- DOW-Loop
 - Also known as Whitlock DO-Loop or Dorfman-Whitlock DO-Loop

Advantages

- Moves DATA step SET statement inside of a DO-Loop
 - Gives complete control over retention of variable values and population of Program Data Vector (PDV)
 - Requires just one pass through the dataset

Resulting Code

```
data out_ct (drop=patct denom rxgrp);  
  array patcts (*) patct1-patct2;  
  array denoms (*) denom1-denom2;  
  
  do until (last.listnum or eof);  
    set out_ct1 end=eof;  
    by analytec listnum;  
  
    patcts(rxgrp) = patct;  
    denoms(rxgrp) = denom;  
  end;  
  
  output;  
run;
```

How It Works (Part 1)

- Initialization
 - DATA statement initializes PDV
 - ARRAY statements are only executed for the first value of ANALYTEC/LISTNUM

How It Works (Part 2)

- Individual record processing
 - DO-loop executes over all records with the same ANALYTEC/LISTNUM value
 - SET statement populates PDV
 - Appropriate array elements are assigned values based on the value of RXGRP
 - Note that array elements are retained until all BY LISTNUM processing is done

How It Works (Part 3)

- Output
 - OUTPUT statement is executed after the DO-loop completes
 - Output dataset is 1 record per ANALYTEC/LISTNUM

Conclusion

- The DOW-Loop is an extremely powerful technique
 - Gives the programmer more control over input and retention of variables
 - Reduces the number of passes through a dataset required by the program
 - Fewer passes → faster programs!

Contact Information

Nancy Brucken
STATPROBE, Inc.
5430 Data Court, Suite 200
Ann Arbor, MI 48108
(734) 769-5000 x1231
Nancy.Brucken@statprobe.com