Selected Topics from the DATA Step

by Ian Whitlock
Ian.Whitlock@comcast.com

Topics

- DATA step
  - Named INPUT
  - UPDATE vs MERGE
  - Simple rules for MERGE

Named INPUT

- Named PUT is important in debugging

```
put var= ;
```

- So what happened to named INPUT?

```
input var= ;
```


Named INPUT

```
data w ;
   length name $ 20 ;
   input x= y= name=;
cards ;
x=1 y=2 name=Jane McHooly
y=4
;
```

- No blank between variable name and equal sign

Named INPUT - Features

```
data w ;
   length id 8 name $ 20 ;
   input id= ;
cards ;
id= 123 name= Martha Done
;
```

- Looks for any variable in the PDV
- Uninitialized message in version 8.2

Named INPUT - Features

```
data w ( drop = end ) ;
   length x y date 8 end $ 1 ;
   informat date date9. ;
   input ( _all_ ) ( = ) ;
cards ;
y=2 /        continue next line
date=22Sep2003 /
end=*        easy to spot
;
```

Named INPUT

- Superb for constructing tailored test data
  - Specify only important variables for test
  - Readable for complex data
  - Add NOTE variable giving desired result

7

UPDATE Statement

- Master - one record per ID
- Trans - multiple records per ID possible
- Standard missing means no change

```
data master ;
    update master trans ;
    by id ;
run ;
```

8

UPDATE Statement

Master file

ID	FirstName	LastName	MStatus
00001	Jack	Arthur	S
00002	Mary	Wooden	S
00003	Bill	Bull	M

9

Creating Transactions

```
data stdtrans ;
   length ID $5
          FirstName  $20
          LastName $ 30
          MStatus $ 1
   ;
   input ( _all_ ) ( = ) ;
cards ;
id=00001 firstname=John
id=00002 lastname=Smith
id=00002 mstatus=M
;
```

10

The Update

```
data new_master ;
   update old_master stdtrans;
   by id ;
run ;
```

11

After Update

Master file

ID	FirstName	LastName	MStatus
00001	John	Arthur	S
00002	Mary	Smith	M
00003	Bill	Bull	M

12

Excel Transactions

ID	Variable	Value
00001	FirstName	John
00002	LastName	Smith
00002	MStatus	M

13

Creating Trans

```
filename temp temp ;
data _null_ ;
   set xls.trans ;
   file temp ;
   put id= variable +(-1)"=" value
  ;
run ;
```

14

External Temp File

```
id=00001 FirstName=John
id=00002 LastName=Smith
id=00002 MStatus=M
```

15

Simple Rules for Stable MERGE Programs

- Always use a BY statement (at least set option MERGENOBY=ERROR)
- BY variables must be in common
- Corresponding BY variables must have same length (Bob Virgile) and type
  http://www.nesug.org/Proceedings/nesug03/at/at005.pdf
- No other variables in common
- Do not change higher level information without reserving names

16

Simple Rules for Stable MERGE Programs

- Any desire to break a rule is:
  - A mistake, or
  - Indicates MERGE is wrong tool
- Alternatives:
  - SET/BY for interleaving
  - UPDATE/BY for updating

17

Summary

- Named INPUT
  - Good for creating test data
  - Good for transactions with UPDATE
- UPDATE vs MERGE
  - Simple rules to control good merge
    - ✔ Always use BY statement
    - ✔ Common keys and common attributes
    - ✔ No other variables in common
    - ✔ No changes to higher level data

18

Notes

SESUG 2003

A Sampler of Code Techniques

Speaker Information

Ian Whitlock
29 Lonsdale Lane
Kennett Square, PA, 19348-2045

Ian.Whitlock@comcast.com