

Missing Secrets

Ian Whitlock

ian.whitlock@comcast.com

The DATA step

- Implied loop
- Driven by input
- Variables
 - Initialized once
 - May be reset by the system
 - May be changed by user code

2

The Problem

- Some variables set to missing
 - When?
 - What are the rules?

3

Variable Classification

- Unnamed variables
- Variables from SAS data sets
- Automatic variables
- All other variables

4

Unnamed Variables

- Initialized to missing
 - unless explicitly initialized
- Always retained
 - i.e. requires user code to change
- Created in temporary arrays
 - `array temp (100000) _temporary_ ;`

5

Automatic Variables

- Always retained
- Reset by system when appropriate
- May be assigned by user
- Examples:
 - `_n_`
 - `last.var`
 - `in=`

6

All Other Variables

- Set to missing at the top of each implied DATA step loop by default
- Values not reset when in explicit RETAIN statement
- Initialized to missing unless explicitly initialized in RETAIN statement

7

Variables from SAS Data sets

- Always retained
 - May be reset by system or user
 - Initialized to missing at the top of the first implied DATA step loop
- Exception:
- Can be explicitly initialized in a RETAIN statement

8

Variables from SAS Data sets

- When does SAS set to missing?
 - At top of first iteration of DATA step
 - At SET statement when switching to a data set that does not contain the variable
 - At MERGE statement when beginning a BY-group and the variable is in one of data sets of the merge, but not read in this BY-group

9

Variables from SAS Data sets

- SET statement

```
data _null_ ;
  put x= ;
  set w1 (keep=x) w2 (keep=y) ;
  put x= ;
  x = 6 ;
run ;
```

 - X is missing when first record is read from W2 and 6 thereafter

10

Variables from SAS Data sets

- MERGE statement

```
data _null_ ;
  merge w1 ( keep = key x )
        w2 ( keep = key y ) ;
  by key ; put x= ;
run ;
```

 - X is missing on first record of BY-group whenever W1 is not in the BY-Group

11

Conclusion

- The rules are difficult to state precisely without the variable classification
- To understand:
 - Distinguish setting to missing at the top of the loop from setting to missing when not read
 - SAS records are read only once

12

Notes

SAS is a registered trademark the SAS Institute Inc in the USA and other countries.

® indicates USA Registration.

Taken from my response to a SAS-L question
Presented as a poster with Robin Jones SESUG 2002 -
"Missing Secrets"

13

Speaker Information

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

Ian.Whitlock@comcast.com

14