

Paper CC14

Sorting out the Possibilities Using Retain

Tommy Retzlaff, Quintiles, Morrisville, NC

Cara Woodruff, Quintiles, Morrisville, NC

ABSTRACT

Controlling the order of your data along with some simple assumptions can make creating analysis flags much easier for the beginner to intermediate programmer. This example will look at conquering the multiple-choice scenario of stacked medication interruption diary data. We are interested in categorizing this data into one of three separate possible results by using one temporary variable and with only one pass through the data. This will allow us to define whether the study medication was 1) NEVER INTERRUPTED, 2) INTERRUPTED, or 3) UNKNOWN whether it was interrupted.

INTRODUCTION

The idea of creating analysis flags across multiple records can be a difficult concept for many programmers. Using the sample data below, I hope to show that with the control of the SAS® SORT procedure, RETAIN statement, and a simple assumption we can easily obtain the multiple-choice, prioritized result with one pass through the data step. The sample code will be used to determine for a patient whether 1) study medication was NEVER INTERRUPTED, 2) study medication was INTERRUPTED, or 3) study medication interruption is UNKNOWN. The sample test data below contains stacked records of medication diary data for a multiple visit study for each patient. The variable INTRUPT has three possible values: N=Not interrupted, Y=Interrupted, Missing=Unknown or Not Recorded.

Sample Data:

Obs	SUBJID	VISITN	INTRUPT
1	4	1	N
2	4	2	N
3	4	3	N
4	4	4	N
5	4	5	N
6	4	6	N
7	4	7	N
8	4	8	N
9	8	1	N
10	8	2	N
11	8	3	Y
12	8	4	N
13	8	5	N
14	8	6	N
15	8	7	N
16	8	8	N
17	12	1	N
18	12	2	N
19	12	3	N
20	12	4	N
21	12	5	N
22	12	6	N
23	12	7	N
24	12	8	N

METHODOLOGY

By using the SAS SORT procedure, the order of the data is controlled so that all diary information can be considered chronologically through one pass of the data and the categorical result indicator (INT_CAT) can be created on the last available record. Use of the RETAIN statement allows us to make an assumption from the start of each patient using a temporary flag (INT_FLAG). The temporary flag will record the different scenarios as they occur. In this case, we will make the simple assumption that medication is NEVER INTERRUPTED (INTRUPT="N") by setting the value of our temporary flag to 1 and look to "break" it with the other possible responses (INTRUPT="Y" or " "). Values for INTERRUPTED (INTRUPT="Y") will take the highest priority followed by UNKNOWN (INTRUPT=" ").

SAS Code:

```
proc sort data=drug;
  by subjid visitn;
run;
```

```
*****
***** Create INT_CAT *****
**** 1 = Never Interrupted *****
**** 2 = Interrupted *****
**** 3 = Unknown *****
*****
```

```
data drug;
  set drug;
  by subjid visitn;
```

```
*****
**** create INT_CAT ****
*****
```

```
*** Set temporary flag to our assumption - 1=NEVER ***;
retain int_flag 1;
if first.subjid then int_flag=1;
```

```
*** Highest priority is 2=INTERRUPTED ***;
if intrupt="Y" then int_flag=2;
else if intrupt=" " and int_flag^=2 then int_flag=3;
```

```
*** Set categorical result indicator on last record for each patient ***;
if last.subjid then int_cat=int_flag;
```

```
run;
```

Results:

Obs	SUBJID	VISITN	INTRUPT	int_flag	int_cat
1	4	1	N	1	.
2	4	2	N	1	.
3	4	3	N	1	.
4	4	4	N	1	.
5	4	5	N	1	.
6	4	6	N	1	.
7	4	7	N	1	.
8	4	8		3	3
9	8	1	N	1	.
10	8	2	N	1	.
11	8	3	Y	2	.
12	8	4	N	2	.
13	8	5	N	2	.
14	8	6	N	2	.
15	8	7	N	2	.
16	8	8	N	2	2
17	12	1	N	1	.
18	12	2	N	1	.
19	12	3	N	1	.
20	12	4	N	1	.
21	12	5	N	1	.
22	12	6	N	1	.
23	12	7	N	1	.
24	12	8	N	1	1

CONCLUSION

After passing through the data step, we can see that each of the 3 possible scenarios has been successfully assigned. The use of the assumption that always or never (100% scenario) occurs allows us to easily assign values that occur sometimes. The "sometime" values can then be prioritized to a certain order. This eliminates the necessity of passing through the data several times to find out what occurred. (Note: the temporary variable would normally be dropped.)

Contact Information

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

Tommy Retzlaff

Manager, Statistical Programming

Quintiles, Research Triangle Park

Office: (919) 998-1470

Fax: (919) 998-2382

Email: tommy.retzlaff@quintiles.com

Cara Woodruff

Associate Manager, Statistical Programming

Quintiles, Research Triangle Park

Office: (919) 998-7166

Fax: (919) 998-2382

Email: cara.woodruff@quintiles.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.