

Does NODUPKEY Select the First Record in a By Group?

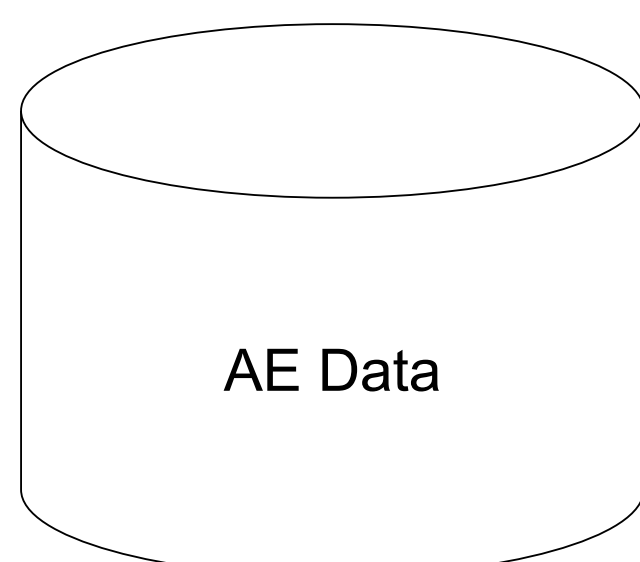
David Franklin, TheProgrammersCabin.com, Litchfield, NH

INTRODUCTION

For some, the NODUPKEY option in the SORT procedure always selects the first observation in a group of variables. But does it? Read on and find out!

FIRST, SOME DATA

```
data ae0;
  input ptnum $6. aeseq event $20.;
cards;
001002 1 HEADACHE
001001 1 FEVER
001001 2 HEADACHE
001003 1 NAUSEA
001003 4 DIARRHEOA
001003 2 VOMITING
001004 1 DIARRHEOA
001001 3 DIARRHEOA
001002 2 DIARRHEOA
001004 2 HEADACHE
001003 3 FEVER
;
```



First, a simple basic SORT procedure call that outputs the unique PTNUM values:

```
proc sort data=ae0 out=ael nodupkey;
  by ptnum;
proc print data=ael;
  title1 "Output Using NODUPKEY";
run;
```

Output Using NODUPKEY

Obs	ptnum	aeseq	event
1	001001	1	FEVER
2	001002	1	HEADACHE
3	001003	1	NAUSEA
4	001004	1	DIARRHEOA

THE TWINS ...

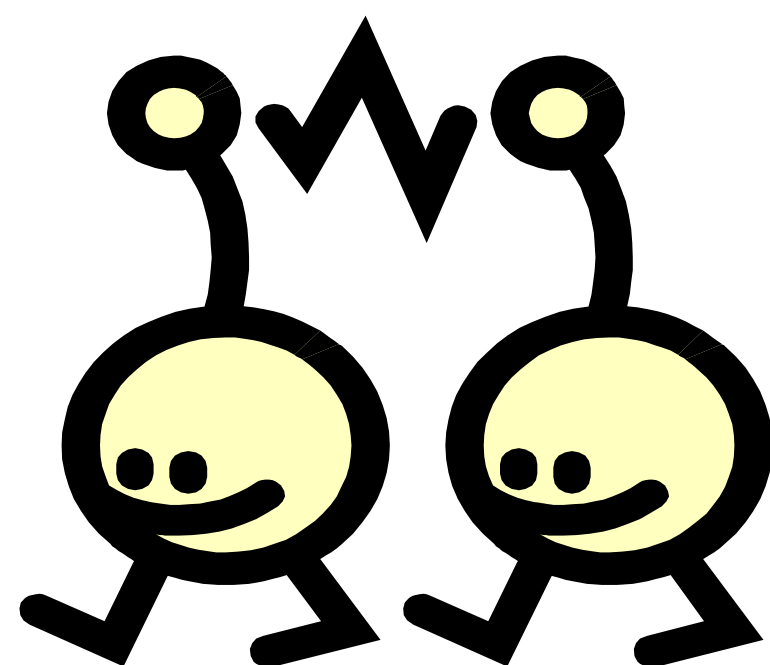
Welcome the EQUALS|NOEQUALS option in the SORT procedure, and the global option SORTEQUAL|NOSORTEQUALS – it is these two options that effect whether the first observation in a group is selected or not. The documentation for the SORTEQUAL|NOSORTEQUALS is given below:

SORTEQUALS

specifies that observations with identical BY variable values are to retain the same relative positions in the output data set as in the input data set.

NOSORTEQUALS

specifies that no resources be used to control the order of observations with identical BY variable values in the output data set.



Their counterparts in the SORT procedure are EQUALS for SORTEQUALS and NOEQUALS for the NOSORTEQUALS. The default value when SAS is shipped is SORTEQUALS.

It is interesting to note that SAS suggests that in order to save resources, use NOSORTEQUALS when you do not need to maintain a specific order of observations with identical BY variable values.

... AND OH WHAT TROUBLE THEY CAN MAKE

Now lets look at what these options can do – first with the EQUALS option used:

```
proc sort data=ae0 out=ael nodupkey equals;
  by ptnum;
proc print data=ael;
  title1 "Output Using NODUPKEY and EQUALS Options";
run;
```

Output Using NODUPKEY and EQUALS Options

Obs	ptnum	aeseq	event
1	001001	1	FEVER
2	001002	1	HEADACHE
3	001003	1	NAUSEA
4	001004	1	DIARRHEOA

This is the output we got previously.

Now lets look at if the EQUALS option was changed to the NOEQUALS option:

```
proc sort data=ae0 out=ael nodupkey noequals;
  by ptnum;
proc print data=ael;
  title1 "Output Using NODUPKEY and NOEQUALS Options";
run;
```

Output Using NODUPKEY and NOEQUALS Options

Obs	ptnum	aeseq	event
1	001001	3	DIARRHEOA
2	001002	1	HEADACHE
3	001003	3	FEVER
4	001004	1	DIARRHEOA

The output now is not the first record in the output sequence, nor is it the first record found for each group, i.e. PTNUM value.

Similar results are attained using the global SORTEQUAL|NOSORTEQUALS options.

One advantage of using the NOEQUALS or NOSORTEQUALS option is that the sort procedure saves resources, which can be valuable when dealing with large datasets.

Using NOEQUALS or NOSORTEQUALS saves resources, valuable when sorting large datasets!

CONCLUSION

So,

“Does the NODUPKEY option in the SORT procedure always select the first observation in a group of variables?”

And the answer is

“Not Really”

but the two options that do effect this are the EQUALS|NOEQUALS options in the SORT procedure, and the global options SORTEQUAL|NOSORTEQUALS.

REFERENCES

SAS Institute Inc. 2006. Base SAS® 9.1.3 Procedures Guide, Second Edition, Volumes 1, 2, 3, and 4. Cary, NC: SAS Institute Inc.

CONTACT INFORMATION

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

Author Name: David Franklin
 Enterprise: TheProgrammersCabin.com
 City, State ZIP: Litchfield, NH 03052
 Work Phone: 603-275-6809
 Email: dfranklin@TheProgrammersCabin.com
 Web: TheProgrammersCabin.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. Other brand and product names are trademarks of their respective companies.

