Checking Out Your Dates **with SAS®**

Christopher Bost

What's in it for you?

- All variables must be checked
  - PROC FREQ [categorical]
  - PROC MEANS [continuous]
- Date variables are **different**
  - PROC FREQ
  - PROC MEANS
  - PROC TABULATE

Sample data set: PREP

ID	DOB	Enrolled	Completed
101	06/02/1995	08/31/2012	09/15/2012
102	05/22/1995	08/30/2012	10/01/2012
103	08/15/1995	09/15/2012	09/30/2012
104	07/07/1995	08/31/2012	.
105	12/14/1994	08/30/2012	09/10/2012
106	01/03/1995	09/04/2012	10/16/2012
107	04/05/1995	09/04/2012	11/01/2012
108	11/11/1994	08/30/2012	11/10/2012
109	01/30/1995	09/04/2012	.
110	04/15/1995	09/04/2012	09/12/2012

Desired output

- Variable name
- Variable label
- Sample size
- Missing values
- Minimum
- Maximum
- Median
- Range

PROC FREQ: Syntax

```
proc freq data=prep;
tables DOB Enrolled Completed;
run;
```


Table of DOB

Date of birth				
DOB	Frequency	Percent	Cumulative Frequency	Cumulative Percent
11/11/1994	1	10.00	1	10.00
12/14/1994	1	10.00	2	20.00
01/03/1995	1	10.00	3	30.00
01/30/1995	1	10.00	4	40.00
04/05/1995	1	10.00	5	50.00
04/15/1995	1	10.00	6	60.00
05/22/1995	1	10.00	7	70.00
06/02/1995	1	10.00	8	80.00
07/07/1995	1	10.00	9	90.00
08/15/1995	1	10.00	10	100.00

Table of COMPLETED

Date finished				
Completed	Frequency	Percent	Cumulative Frequency	Cumulative Percent
09/10/2012	1	12.50	1	12.50
09/12/2012	1	12.50	2	25.00
09/15/2012	1	12.50	3	37.50
09/30/2012	1	12.50	4	50.00
10/01/2012	1	12.50	5	62.50
10/16/2012	1	12.50	6	75.00
11/01/2012	1	12.50	7	87.50
11/10/2012	1	12.50	8	100.00

Frequency Missing = 2

Table of ENROLLED

label → Date started

	Enrolled	Frequency	Percent	Cumulative Frequency	Cumulative Percent
name →					
min →	08/30/2012	3	30.00	3	30.00
	08/31/2012	2	20.00	5	50.00
	09/04/2012	4	40.00	9	90.00
max →	09/15/2012	1	10.00	10	100.00

n

~~median~~

~~range~~

Recommendation

- Use PROC FREQ to check date variables with a **limited number of values**

PROC MEANS: Syntax

```
proc means data=prep
   n nmiss min max median range;
var DOB Enrolled Completed;
run;
```

PROC MEANS output

Variable	Label	N	NMiss	Minimum	Maximum	Median	Range
DOB	Date of birth	10	0	12733.00	13010.00	12883.00	277.0000000
Enrolled	Date started	10	0	19235.00	19251.00	19238.00	16.0000000
Completed	Date finished	8	2	19246.00	19307.00	19266.50	61.0000000

Recommendation

- Do **not** use PROC MEANS output to check date variables

PROC TABULATE: Syntax

```
proc tabulate data=prep;
var    DOB Enrolled Completed;
table  DOB Enrolled Completed,
       n nmiss
       (min max median)*f=mmddyy10.
       range;
run;
```


PROC TABULATE output

	N	NMiss	Min	Max	Median	Range
Date of birth	10	0	11/11/1994	08/15/1995	04/10/1995	277.00
Date started	10	0	08/30/2012	09/15/2012	09/02/2012	16.00
Date finished	8	2	09/10/2012	11/10/2012	09/30/2012	61.00

Recommendation

- Use PROC TABULATE to check date variables

PROC MEANS ODS output data set

```
ods output summary=odsmeans;
proc means data=prep
  n nmiss min max median range;
var DOB Enrolled Completed;
run;
```

Obs	VName_DOB	Label_DOB	DOB_N	…	DOB_Min	DOB_Max	VName_Enrolled
1	DOB	Date of birth	…	…	…	13010	Enrolled

Obs	Label_Enrolled	Enrolled_N	…	…	Enrolled_Min	Enrolled_Max	VName_Completed
1	Date started	10	…	…	19235	19251	Completed

Obs	Label_Completed	Completed_N	…	…	Completed_Min	Completed_Max
1	Date finished	…	…	…	19246	19307

PROC MEANS revisited

SAS 9.3

```
ods output summary=stacked;
proc means data=prep
           n nmiss min max median range
           stackods;
var DOB Enrolled Completed;
run;
```


PRINT of “stacked” data set

Variable	Label	N	NMiss	Min	Max	Median	Range
DOB	Date of birth	10	0	12733	13010	12883	277.000000
Enrolled	Date started	10	0	19235	19251	19238	16.000000
Completed	Date finished	8	2	19246	19307	19267	61.000000

Specify the format to use

```
proc print data=stacked noobs;
format min max median mmddyy10.
       range 3.0;
run;
```


Formatted “stacked” data set

Variable	Label	N	NMiss	Min	Max	Median	Range
DOB	Date of birth	10	0	11/11/1994	08/15/1995	04/10/1995	277
Enrolled	Date started	10	0	08/30/2012	09/15/2012	09/02/2012	16
Completed	Date finished	8	2	09/10/2012	11/10/2012	09/30/2012	61

Recommendation

- Use the PROC MEANS **stacked output data set** to check date variables

Score card

	FREQ	MEANS	TABULATE	STACKODS
Name	■	■	□	■
Label	■	■	□	■
N	■	■	■	■
NMiss	■	■	■	■
Min	■		■	■
Max	■		■	■
Median			■	■
Range		■	■	■

Conclusion

- PROC FREQ, MEANS, and TABULATE can be used to check date variables
- FREQ and MEANS have significant limitations
- TABULATE does almost everything
- PROC MEANS stacked output data set is best

Feedback

Questions and comments are valued and encouraged.

christopher.bost@mdrc.org

chrisbost@gmail.com