

Y2K Windowed Two Digit Years and When to Use YEARCUTOFF

Clarence Wm. Jackson, CQA

Abstract

The Y2K situation has caused Information Services Professionals world wide to react to using two digit-years in implied representation of the century. Raw data files in general did not include the full four digit representation of the year, programs reading the data would have defective results. Several program and raw data remediation methods have been developed and deployed to correct the problem. Such methods include full expansion of the year data to include the century, and window logic to work around the two digit year data. Full expansion, while the best method that provides the best solutions, is an expensive process, so most shops have used window logic to provide the solution. Window logic allows the raw data to remain in two digit form. Raw data from processes using window logic requires the SAS programmer to use the YEARCUTOFF= option to reliably read the data. This paper defines when and how to use the YEARCUTOFF= option, when it's not required, and how to know the difference.

Introduction

SAS has data handling features that must be understood before you proceed to use dates in the SAS System. The use of SAS dates isn't new, but the addition of the Y2K situation has cause some changes in the way everyone should think about dates. Dates in computers are really sets of paired numeric data, store in a specific format depending on the operating system. If you have worked with dates, and have been a part of non SAS Y2K remediation projects, you may have seen dates in many different formats. The two digit year dates that may have slipped by Y2K remediation through a windowing process are date data that will draw increased scrutiny in the future.

The reading of two digit year dates will trigger the YEARCUTOFF= option. When SAS read raw data, the informats define the representation of the variables for SAS to process. When this data is date data and a date informat is used to define the data, SAS will convert the data into SAS date serial, a numeric number based on "Jan 1, 1960", with Jan 1, 1960 having a numeric value of "0". SAS Version 6 ships with the YEARCUTOFF set to equal "1900". What do you want your SAS program to do with two digit years on and after Jan 1, 2000?

Working With Two Digit Years in Raw Data

The use of "Windowing Logic" as the major remediation method in addressing the Y2K problem has left more problems. If you want your SAS program to read two digit years on and after Jan 1, 2000, you're forced to do a little background work on the date data before being assured that you have correctly processed the data.

If you are not the owner/maintainer of the data, you need to understand the dates that are on the raw files. You should ask question regarding how the dates are handled. If you don't know,

you could be making bad assumptions regarding the raw data. If the dates were expanded, your problems are solved in this regard. If the year date data is in a two digit format, you need to know the basic characteristics about the data. You need to ask the following

- Does the year date data values span over 99 years?
 - If so, what and where is the century indicator?
- Was windowing logic used?
 - If so, what pivot year was used for the date data element/field?
 - Was the pivot year fixed or sliding?
 - If it was fixed, what value was used?
 - If sliding, what algorithm was used to define the value?

Windowing has brought on another set of concerns regarding dates, and it's time to understand how to deal with it. The problems of misunderstanding any of the above attributes in working with two digit years in raw data can pose a serious contamination problem if not watched carefully. Carelessly handling dates can no longer be afforded. What you don't know you don't know. Ask!!!

What SAS Can Do With Two Digit Years in Raw Data

The way that you should handle date data with these characteristics will be to understand

- how SAS handles dates
- what the YEARCUTOFF= option is and what it does
- how you can use the YEARCUTOFF= option to minimize errors in reading two digit years
- how to find out from your raw data what value to set your program's YEARCUTOFF= option
- why you need to change your YEARCUTOFF= option from its shipped value for Version 6
- when you need to set your YEARCUTOFF= option

What SAS will do when it encounters two digit year data is to refer to the value of the current YEARCUTOFF= option to 'window logic' the data. The YEARCUTOFF= option defines the beginning of the 100 year window in which SAS will interpret the date from two digit informats, formats, and functions.

So if the value for the YEARCUTOFF= option is set at "1900", then a two digit date variable "twodigit" would have the value "Jan 1, 1900", not "2000". If the value of "twodigit" was meant to be "Jan 1, 2000", this presents a problem, and also illustrates the Y2K problem of incorrect and inaccurate date processing.

```
Data Y2Kdate;
Infile Cards;
Input @1 twodigit mmdyy6.;
Cards;
010100
;
proc print;
run;
```

So what should be done? Set the YEARCUTOFF= option to the value that's best for the raw data. If "YEARCUTOFF=1901", I would get the correct value for "twodigit". For this data element's value, setting "YEARCUTOFF=1901" works. But for multiple values that may be on the files, something more is needed, and a little research on the data as well.

How do you know what the oldest year date data is on raw files? You find out using SAS stuff like PROC FREQ, or you can use a DATA step before to subset your date data, and do something like this:

```
Data whosoldr;
Infile Cards;
Input @1 twodigit mmdyy6.;
If year(twodigit) LE '010120'D then b1920='Yes';
    Else b1920='No ';
Cards;
010100
012283
.
```

Title "Years Before and After Proposed YearCutOff= Value of 1920";

```
Proc Freq;
Tables b1920;
Run;
```

Title "Frequency of Years Before Proposed YearCutOff= Value of 1920";

```
Proc Freq;
Where (b1920="Yes");
Tables twodigit;
Run;
```

The above code will produce two nice reports, the number of records that are before and after the proposed YEARCUTOFF= date, and the values for years before the date. Now I know if I'm going to have a problem with my date, because it's been measured. If there is a program that needs to read the data, I can set the program's YearCutOff= to the appropriate value.

NOTE TO SAS CONSULTANTS AND INHOUSE SAS INSTALLATION REPS – PLEASE READ!!!

There is another need that must be addressed, and that is the use of SAS in production systems where the clients of the software is not aware of what Y2K and YEARCUTOFF='1900' will do to their data. By now, a Y2K inventory should have been done that should alert the SAS Site Reps and Consultants regarding the use of dates by their clients. The YEARCUTOFF= should be set to something other than '1900' before December 31, 1999.

For those that don't know or care what a "YearCutOff=" is, it's up to those that administer SAS for the enterprise to set the value of the option YearCutOff<>1900. Why! See example above. For those in your "care" that haven't taken the time to review their programs, you need to set the value to something other than "1900", or you will be getting calls from folks in a panic.

You need to set the value to something other than "1900" before Jan 1, 2000 for the default within you shop. This should be set at the SAS configuration level or the procedure level (for mainframes).

Conclusion

The Y2K problems will be with us long after Jan 1, 2000 is over. Be careful, as there are a mix of different data out there claiming to be Y2K compliant, when it's really Y2K windowed. If it's not expanded, it's not Y2K compliant data. When you get the data without the application logic or explanation regarding how the dates should be processed, you should seek to understand the data; it's attributes, and conditions for using the data. Windowed data will require you to pay close attention to the logic used, and it can be a "fun" experience. Learning to use SAS dates and the YEARCUTOFF= option can go a long ways toward helping you get the data through cleanly. Thank you for your time.

REFERENCES

SAS Institute Web Site on Year 2000, www.sas.com/y2k

SAS Language Guide, "SAS DATES", Version 6, SAS Institute Inc

Jackson, Clarence Wm., *SAS and the Year 2000 Problem*, Proceedings of the Sixth Annual South-Central Regional SAS Users' Group Conference, SCSUG '96

Jackson, Clarence Wm., *Using SAS to Help Manage Data*, Proceedings of the Fifth Annual South-Central Regional SAS Users' Group Conference, SCSUG '95

The author can be contacted via email

Clarence Wm. Jackson, CQA
CJac@compuserve.com
<http://ourworld.compserve.com/homepages/CJac>
cljacks@ci.dallas.tx.us (City of Dallas)