When a Period Really is Not a Period and How to Fix It

ABSTRACT
The other day I was working on a clinical study for a client in which I needed to import and convert a Microsoft Excel spreadsheet into a SAS® Version 9 dataset. The spreadsheet contained many variable columns that contained actual numeric values, a period (‘.’) to denote one level of missing, and an empty space (‘ ’) to denote a different level of missing. In order to accommodate the aforementioned columns that contained a combination of numbers and characters, I imported the columns as character variables. I noticed after running my tried and true DDE code that imports Excel into SAS that all of my periods ‘.’ were now in the SAS dataset as empty spaces ‘ ’. Puzzled, but not deterred, I tried several methods to create a correct SAS dataset with periods and spaces. This paper will explore the different steps I tried and the final result used to create an equivalent SAS dataset from the original Excel spreadsheet.

SURPRISE! SURPRISE!

Our company was working on a marketing study for a client. We had been forwarded some data that was entered on Microsoft Excel spreadsheets and were asked to prepare some SAS analyses. We have developed a SAS macro tool that has been in production for a couple of years that utilizes DDE (Dynamic Data Exchange) to convert the Excel spreadsheet into SAS Version 9 datasets. An excel spreadsheet that is similar to the ones used in our study is shown in Figure 1.

Figure 1: Sample Excel Data

I pulled the macro code out of our library of macros, ran it, and Voila!, it produced several SAS datasets that were to be used for statistical analyses. I told our statistician that the data were ready and I went about my day doing other data management tasks. The problem began a couple days later when our statistician started to try to make tables and found some discrepancies in one particular dataset. You can see the initial DDE transferred SAS data in Figure 2. It turned out that the key measures that were imported as numeric variables actually were supposed to have two values of missing that denoted separate conditions. The data entry staff entered missing values of a blank to denote the first missing condition (subject missed the visit measurement) and a period to denote the second missing condition (the key measure could not be measured). I thought, no problem, I’ll modify my DDE code to import the scoring variables as character, thus maintaining the score, blank, and period values.
GOLLY, THAT PERIOD SURE LOOKS LIKE A BLANK

After I modified the DDE code, ran it again, and reviewed the new SAS dataset, much to my chagrin, the new scoring character variables only had spaces and no periods. My first thought was maybe data entry did not properly enter periods and blanks. I went back to the source Excel workbook and there clearly were blanks and periods in the key measurement fields. Next, I thought that perhaps this was a problem with Excel. I looked at the source workbook and examined the format of the scoring columns and noticed that they were the General format. I was certain that the Excel formatting was the problem, so I selected all scoring columns, converted them to a Text format, and saved the file. I reran the Excel to SAS macro, checked the log and found no error messages, but opened the resultant SAS dataset and still like the game rock, paper, scissors, blank overruled period and my dataset had no periods (Figure 3).

At this point, I was out of ideas of how to fix this issue. I then called SAS Tech Support, who logged my issue and said that they would get back to me after researching the issue, but initially seemed stumped regarding a solution. In the meantime, I decided that perhaps the answer might be in SAS help.

THE FINAL ANSWER

Using SAS help can be an adventure in itself as you enter a search term and help presents you with a multitude of choices, but there seems to be no relevance to the results as in other search engines. After combing through the results, two results seemed promising: Using Dynamic Data Exchange under Windows: DDE Examples and Using Dynamic Data Exchange under Windows: DDE Syntax within SAS. The first clue in the help section under the result: Using Dynamic Data Exchange under Windows: DDE Examples stated. "CAUTION, use caution when using DDE with data values that are blank or missing. For sample code, See Reading Missing Data." I knew I was on to something when I got to the Reading Missing Data help result, as there was sample code and the paragraph, "The INFORMAT statement forces the DATA step to use modified list input, which is crucial to this example. If you do not use the modified list input, you receive incorrect results. The necessity of using..."
modified list input is not DDE specific. You would need it even if you were using data in a CARDS statement, whether your data were blank- or comma delimited*. These statements combined with the sample code gave me the answer that I needed. It displayed code showing an INFORMAT statement coupled with the use of the $CHARx Informat. Now, I needed to test whether this snippet of code would produce the results that I desired. I opened up the Excel workbook, opened up our SAS macro tool code, added the INFORMAT statement to the affected key measurement fields, held my breath, and ran the code. I looked in the log, noticed that there were no errors, and then opened the SAS analysis dataset. Wow! There were periods and spaces in the key analysis variables. I let our statistician know the data were fixed and she continued to proceed with the analyses. The final dataset can be seen in Figure 4.

Figure 4: Final Excel DDE Imported SAS dataset - All params are character and have missing values denoted as spaces and periods

POST-MORTEM

The next day, I received a call from SAS Tech Support. They had found a solution to my problem. Essentially, it was the same response as the solution that I discovered. The Tech person stated that they had been working in SAS a long time and had never thought that this issue was a problem. They added that they were surprised that there were no direct NOTES on the support.sas.com website relating to this issue. They stated that they would strongly urge that a NOTE be created regarding this issue.

CONCLUSION

The purpose of this paper was to demonstrate how to resolve a little known issue with DDE between Microsoft Excel and SAS regarding the importing of a period as a missing value. It explains the process to discover the answer and describes the code used to fix this issue. SAS Code used to show the developmental and final source code is included at the end of this paper.

REFERENCES


AKNOWLEDGMENTS

I would like to thank my boss, Howard Proskin for giving me the time to work on this paper. I would like to thank my colleague programmers: Bill, Adam, and Jillian for their collaborative influence and suggestions that improve my SAS programming.
APPENDIX (SOURCE)

** Title: When a Period Really is not a Period and How to Fix It.;  
** Sample SAS Code.;

***First DDE conversion try - Param1 to Param10 as numbers.;

filename bodsys dde 'excel|sheet1!r2c1:r17c11' ;

data bodysys;
  infile bodsys dlm='09'x notab dsd missover;
  input patid param1-param10 ;
run;

***Second DDE conversion try - Param1 to Param10 as character. Unfortunately, no distinction between missing blank and period values.;

filename bodsys1 dde 'excel|sheet1!r2c1:r17c11' ;

data bodysys1;
  retain patid;
  infile bodsys1 dlm='09'x notab dsd missover;
  length param1-param10 $2;
  input patid param1-param10 $;
run;

***Final DDE conversion try - Param1 to Param10 as character. Finally, a distinction between missing blank and period values was made by using the $char10. informat.;

filename bodsys2 dde 'excel|sheet1!r2c1:r17c11' ;

data bodysys2;
  infile bodsys2 dlm='09'x notab dsd missover;
  informat param1-param10 $char10. ;
  input patid param1-param10 $ ;
run;

CONTACT INFORMATION

William Csont, MPA
Howard M. Proskin & Associates, Inc.
300 Red Creek Drive, Suite 220
Rochester, N.Y. 14623
Work Phone: (585) 359-2420
Fax: (585) 359-0465
E-Mail: wcsont@hmproskin.com
Web: www.hmproskin.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.