Paper 89-26

Format Challenges with PROC REPORT

Lisa M. Schneider, Covance Inc., Princeton, NJ

ABSTRACT

PROC REPORT is a very useful tool in the production of tables, particularly if the data to be presented in the column contain the same kind of information for all rows in the table. A more challenging situation arises when different types of data (i.e., discrete and continuous variables) are used in the same column to represent dissimilar information that often needs to be differentially displayed. If the report also requires the use of a classification variable to define columns of the report, more complications can occur.

The use of COMPUTE statements and an ACROSS variable to handle these conditions is explored. COMPUTE statements conditionally assign different formats to the various data presented in the same column based on the values of variables used to distinguish the rows in the table. An ACROSS variable is used to define the classification structure for the columns. Understanding that the REPORT procedure actually transposes the input data set when an ACROSS variable is used, and thus changes the names of the variables, is crucial to the construction of COMPUTE statements. The usefulness of the OUT= option of the procedure is discussed.

INTRODUCTION

Figure 1 is from a table that displays two different kinds of variables (one continuous and one discrete) as well as different kinds of information to be displayed for each variable type. In addition, there are two groups across the page. Each variable has two columns for each group in which the actual data values are displayed. The example shows that each of the two columns displaying data values within each group has four separate formats used to present the information required in the different rows.

The purpose of this paper is to demonstrate the use of a COMPUTED variable in PROC REPORT to assign different formats within a single column appropriate for the information displayed in each row. These columns of data will be repeated for two levels of a classification variable with the use of an ACROSS variable.

STRUCTURE OF THE INPUT DATA SET

PROC REPORT can easily display data for different levels of a classification variable across the page. The data to be presented in the report for different levels of any classification variable can be separate observations in the input data set. It is not a requirement that all the data values for an actual row in the report be on the same observation of the input data set. The input data set to PROC REPORT can thus be vertical (or normalized) instead of horizontal.

For the example above, the input data with the statistics (SEQ=1 through 5) for the continuous variable AGE (PRNTORD=1) for the two groups (CLASSVAR=1 or 2) and their associated two columns of data (LEFTCOL and RGHTCOL) are given in Figure 2. The levels (also SEQ=1 through 5) of the discrete variable RACE (PRNTORD=2) and their data are also included.

ACROSS VARIABLES

The ACROSS variable in PROC REPORT takes a vertical input data set and transposes it horizontally to display the separate observations across the page. Each column “under” an ACROSS variable is determined by the value of the ACROSS variable (in this example, CLASSVAR) as well as the value of another variable (in this example, either LEFTCOL or RGHTCOL). Thus, two variables collectively define what data appear in the column.

After the transpose has taken place, the variables involved in the ACROSS process no longer have their original names. This is because each column is a combination of two variables and each level of the ACROSS variable will generate new columns as the data set becomes horizontal. Given that there has to be a way to directly reference any given column, PROC REPORT renames these new columns as _Cx_ where x is the number of the new column that the combination of variables represents.

The columns are sequentially numbered starting with the first variable in the COLUMN statement and increment by one as each new variable (excluding the ACROSS variable) appears in the COLUMN statement. The transpose also produces new columns for each level of the ACROSS variable and the variables “under” it. These columns are included in the sequential numbering scheme. In this way, PROC REPORT can uniquely reference each column required in the actual report. The columns defined by an ACROSS variable and another variable can be referenced by a single variable name.

THE OUT=SAS-DATA-SET OPTION

Although the column names after the transpose can be determined as explained above, it is sometimes much easier to just output the data set created by PROC REPORT and look at the column names. To do this use the OUT=*SAS-data-set* option on the PROC REPORT statement. The output data set can then be printed and examined to see how the columns have been named.

Continuing with the example above, a simple PROC REPORT (Figure 3) will produce a report without the special formatting of each row. Outputting and printing the data set created shows the names of the new columns after the transpose from the ACROSS variable has been applied.

Details about the PROC REPORT code

- The comma after CLASSVAR (the ACROSS variable) and the parentheses surrounding LEFTCOL and RGHTCOL indicate to PROC REPORT that the combination of those variables collectively will determine the contents of the columns. This is the usual setup for an ACROSS variable and the variables “under” it.
- When an ACROSS variable is used, the variables PRNTORD and SEQ must be defined as GROUP so that the summary process of GROUP collapses the observations into a single row in the report (as opposed to the detail records that would be produced if these variables were defined as ORDER).
- When a DISPLAY and an ACROSS variable share a column, the report must also contain another variable that is not in the same column. Also, an ACROSS variable without a statistic or an analysis variable associated with it will produce an error. (Since LEFTCOL and RGHTCOL are being used as DISPLAY variables in this example, there is no other numeric variable to use as ANALYSIS.) The variable FOOLRPT is used to resolve these issues. This variable is not on the input data set and is by default an ANALYSIS variable. PROC REPORT will create the variable and use it as the report is generated.

Figure 4 presents the OUT=OUTRPT data set. As can be seen, the variable names for the columns involved in the ACROSS transpose process are _C3_ and _C4_, for the LEFTCOL and RGHTCOL variables respectively, for CLASSVAR=1 observations and _C5_ and _C6_ for CLASSVAR=2 observations.

PRODUCING THE ACTUAL REPORT

The entire code for producing the output shown in Figure 1 is included as Figure 5. Important details will be discussed below.

Details

- The COLUMN statement is expanded to include three new COMPUTED variables. (See the associated DEFINE statements.)
  - ♦ LEFTHEAD contains the formatted text to indicate what data is displayed in the row. The values are the formatted versions of the combination of PRNTORD and SEQ.
  - ♦ NEWLEFT and NEWRGHT are the two differentially formatted versions of the variables LEFTCOL and RGHTCOL, respectively, for each of the two levels of the classification variable CLASSVAR.

- The DEFINE statements for PRNTORD and SEQ have their usage as GROUP. In addition, these variables are not printed. They are used to sort and group the observations accordingly. In combination with the ACROSS variable, GROUP summarizes the separate input observations into a single observation in the report. All observations with the same values for the PRNTORD and SEQ variables are one row in the report. PRNTORD and SEQ are also used to apply different formats to the rows.

- LEFTCOL and RGHTCOL both have the NOPRINT option. These variables are used as the data source for which the COMPUTED variables apply the appropriate formats.

- Two versions of COMPUTE blocks are used in this PROC REPORT. The first type (i.e., the COMPUTE BEFORE blocks for each variable) is used to copy the values of the PRNTORD and SEQ variables into variables (ORDPRNT and SEQUENCE, respectively) used in the COMPUTED variables. This is necessary because PROC REPORT does not have access to the values of the original PRNTORD and SEQ variables as it continues to build the report from left to right.
  - ♦ In the COMPUTE block for PRNTORD, there is also code to generate the first row for each of the variables in the report. The LINE statement creates the header text for each variable by simply formatting the value of PRNTORD. In this report, there is no data displayed on these rows.

- The second type of COMPUTE block is used to create the LEFTHEAD column. This type generates values displayed in the actual report. The value of LEFTHEAD is determined by conditionally assigning a format for the SEQ variable. Based on the value of ORDPRNT (PRNTORD), the appropriate format is used to label the values of SEQ.
  - ♦ Values of SEQ for the AGE variable (ORDPRNT = 1) are formatted using $STATS. The format provides a label for each type of statistic produced for the continuous variable.
  - ♦ Values of SEQ for the RACE variable (ORDPRNT = 2) are formatted using $RACE. Each level of the discrete variable has the appropriate text associated with it.

- The primary purpose of this paper is contained in the COMPUTED variables NEWLEFT and NEWRGHT. For each unique combination of ORDPRNT and SEQUENCE that requires a different format to display the data values in these columns, a separate section of code is needed. This way each row of the report has the correct format assigned to display the data accordingly.
  - ♦ Because an ACROSS variable is also used, the columns are renamed to provide a unique reference to the columns after the transpose is applied. The new columns for the first level of CLASSVAR are _C4_ and _C5_ for the “left” column and _C6_ and _C7_ for the “right” column. The same representations for the second level of CLASSVAR are stored in the variables _C8_ through _C11_. These columns can be seen in Figure 6.
  - ♦ To create the values of the COMPUTED variable for NEWLEFT, the names of the differentially formatted variables are _C5_ and _C9_ for the two CLASSVAR levels. Formatting the actual data values contained in _C4_ and _C8_ (the original LEFTCOL variable) generates these columns. Formatting the original RGHTCOL data columns (_C6_ and _C10_) likewise generates the columns of the NEWRGHT COMPUTED variable (_C7_ and _C11_).
  - ♦ Each combination of ORDPRNT and SEQUENCE that requires a different format is represented in the code. The format is conditionally assigned based on the values of these variables because they denote the actual variable and the statistic to be shown in each row of the report.
  - ♦ The data values are displayed using picture formats. Picture formats are quite useful in PROC REPORT. Numeric values can be written with the extra parentheses, commas, and decimals that are often required in tables by including the characters as part of the PICTURE. Caution should always be used when using PICTURE formats. An inappropriate PICTURE can cause severe misrepresentations of the actual data values and no warning will be issued.
  - ♦ The SPACING=0 in the DEFINE statement for NEWRGHT is needed to move that column immediately next to the NEWLEFT column. This is to give the appearance of the data represented as a single character string.

CONCLUSION

COMPUTED variables can be used to assign different formats to data that appear in the same column. The formats are assigned based on values of other variables that uniquely determine the row in the report. Several data types can be represented in a single column using this technique.

CONTACT INFORMATION

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

Lisa M. Schneider
Covance, Inc.
210 Carnegie Center
Princeton, NJ 08540-6233
Work Phone: 609.419.2160
Fax: 609.520.1754
Email: lisa.schneider@covance.com

Figure 1

```
                                               Group 1              Group 2

Age (years)
     N                                      118                  198
     Mean (S.D.)                             52.9  (10.86)        54.2  (10.48)
     Median                                  53.0                 54.0
     25th Percentile, 75th Percentile        46.0,  60.0          47.0,  61.0
     Minimum, Maximum                        28,    83            31,    81

Ethnic Origin n (%)
     Asian                                    2      (1.7)         1      (0.5)
     Black                                    5      (4.2)         8      (4.0)
     Caucasian                              108     (91.5)       180     (90.1)
     Hispanic                                 2      (1.7)         7      (3.5)
     Other                                    1      (0.8)         2      (1.0)
```

Figure 2

```
                       Input Data Set to PROC REPORT

         PRNTORD    SEQ    CLASSVAR    LEFTCOL    RGHTCOL

            1        1        1        118.000       .
            1        2        1         52.881     10.864
            1        3        1         53.000       .
            1        4        1         46.000     60.000
            1        5        1         28.000     83.000
            1        1        2        198.000       .
            1        2        2         54.222     10.484
            1        3        2         54.000       .
            1        4        2         47.000     61.000
            1        5        2         31.000     81.000
            2        1        1          2.000      1.695
            2        2        1          5.000      4.237
            2        3        1        108.000     91.525
            2        4        1          2.000      1.695
            2        5        1          1.000      0.847
            2        1        2          1.000      0.505
            2        2        2          8.000      4.040
            2        3        2        180.000     90.090
            2        4        2          7.000      3.535
            2        5        2          2.000      1.010
```

Figure 3

```
proc report data=indata nowindows headline headskip split='|' out=outrpt ;

     column prntord seq classvar,(leftcol rghtcol) foolrpt ;

     define prntord  / group   order=internal ;
     define seq      / group   order=internal ;
     define classvar / across  order=internal format=$clas. ;
     define leftcol  / display width=7         format=7.2    ;
     define rghtcol  / display width=7         format=7.2    ;
     define foolrpt  / noprint ;
run ;
```

Figure 4

```
               Output Data Set from PROC REPORT - After ACROSS

PRNTORD    SEQ        _C3_      _C4_        _C5_       _C6_      FOOLRPT    _BREAK_
    1      1       118.000      .        198.000      .          .
    1      2        52.881     10.864     54.222     10.484       .
    1      3        53.000      .         54.000      .          .
    1      4        46.000     60.000     47.000     61.000       .
    1      5        28.000     83.000     31.000     81.000       .
    2      1         2.000      1.695      1.000      0.505       .
    2      2         5.000      4.237      8.000      4.040       .
    2      3       108.000     91.525    180.000     90.090       .
    2      4         2.000      1.695      7.000      3.535       .
    2      5         1.000      0.847      2.000      1.010       .
```

Figure 5

```
proc format ;

     value $clas
                   '1' = 'Group 1'
                   '2' = 'Group 2'
                    ;
      value $varfmt
                   '1' = 'Age (years)'
                   '2' = 'Ethnic Origin n (%)'
                    ;
      value $race
                   '1' = '     Asian'
                   '2' = '     Black'
                   '3' = '     Caucasian'
                   '4' = '     Hispanic'
                   '5' = '     Other'
                    ;
    value $stats
                   '1' = '     N'
                   '2' = '     Mean (S.D.)'
                   '3' = '     Median'
                   '4' = '     25th Percentile, 75th Percentile'
                   '5' = '     Minimum, Maximum'
                    ;
     picture xxz
                 .     = '  0' (noedit)
                 other = '009'
                 ;
     picture xxx
                 .     = ' '
                 other = '009'
                 ;
     picture xxxc
                 .     = ' '
                 other = '009,'
                 ;
     picture xxx_x (round)
                 .     = ' '
                 other = '009.9'
                 ;
     picture xxx_xc (round)
                 .     = ' '
                 other = '009.9,'
                 ;
     picture lxxx_xr (round)
                 .     = ' '
                 other = '0009.9)' (prefix='(')
                 ;
     picture lxx_xxr (round)
                 .     = ' '
                 other = '009.99)' (prefix='(')
                 ;
run ;
```

Figure 5 Continued

```
proc report data=indata missing headline headskip nowindows split='|' out=outrpt ;

     column prntord seq lefthead classvar, (leftcol newleft rghtcol newrght) foolrpt ;

     define prntord  / group    order=internal          noprint                              ;
     define seq      / group    order=internal          noprint                              ;
     define lefthead / computed                width=37                                  ' ' ;
     define classvar / across   order=internal                             format=$clas. ' ' ;
     define leftcol  / display                          noprint                              ;
     define newleft  / computed                width=7          spacing=7                ' ' ;
     define rghtcol  / display                          noprint                              ;
     define newrght  / computed                width=8          spacing=0                ' ' ;
     define foolrpt  /                                  noprint                              ;

     compute before prntord ;
        line @6 prntord $varfmt. ;
        ordprnt = prntord ;
     endcomp ;

     compute before seq ;
        sequence = seq ;
     endcomp ;

     compute lefthead / char length=37 ;
             if (ordprnt eq '1') then lefthead = put(seq, $stats.) ;
        else if (ordprnt eq '2') then lefthead = put(seq, $race. ) ;
     endcomp ;

                        ***  Left Column  *** ;

     compute newleft / char length=7 ;

                        ***  Discrete Variables - Count  *** ;
             if (ordprnt in('2')) then
                 do ;
                     _c5_  = put(_c4_, xxz.) ;
                     _c9_  = put(_c8_, xxz.) ;
                 end ;

                        ***  Continuous Variables - N  *** ;
        else if ((ordprnt in('1')) and (sequence in('1'))) then
                 do ;
                     _c5_  = put(_c4_, xxz.) ;
                     _c9_  = put(_c8_, xxz.) ;
                 end ;

                        ***  AGE - Remaining Statistics  *** ;
        else if (ordprnt in('1')) then
                 do ;
                        ***  AGE - Mean & Median  *** ;
                           if (sequence in('2' '3')) then
                               do ;
                                   _c5_  = put(_c4_, xxx_x.) ;
                                   _c9_  = put(_c8_, xxx_x.) ;
                               end ;

                        ***  AGE - 25th Percentile  *** ;
                      else if (sequence in('4')) then
                               do ;
                                   _c5_  = put(_c4_, xxx_xc.) ;
                                   _c9_  = put(_c8_, xxx_xc.) ;
                               end ;

                        ***  AGE - Minimum  *** ;
                      else if (sequence in('5')) then
                               do ;
                                   _c5_  = put(_c4_, xxxc.) ;
                                   _c9_  = put(_c8_, xxxc.) ;
                               end ;
                 end ;
     endcomp ;
```

Figure 5 Continued

```
                        ***  Right Column  *** ;

     compute newrght / char length=8 ;

                        ***  Discrete Variables - Percent  *** ;
             if (ordprnt in('2')) then
                 do ;
                    _c7_  = put(_c6_, lxxx_xr.) ;
                    _c11_ = put(_c10_,lxxx_xr.) ;
                 end ;

                        ***  Continuous Variables - N & Median  *** ;
        else if ((ordprnt in('1')) and (sequence in('1' '3'))) then
                 do ;
                    _c7_  = ' ' ;
                    _c11_ = ' ' ;
                 end ;

                        ***  AGE - Remaining Statistics  *** ;
        else if (ordprnt in('1')) then
                 do ;
                        ***  AGE - SD  *** ;
                           if (sequence in('2')) then
                               do ;
                                   _c7_  = put(_c6_, lxx_xxr.) ;
                                   _c11_ = put(_c10_,lxx_xxr.) ;
                               end ;

                        ***  AGE - 75th Percentile  *** ;
                      else if (sequence in('4')) then
                               do ;
                                   _c7_  = put(_c6_, xxx_x.) ;
                                   _c11_ = put(_c10_,xxx_x.) ;
                               end ;

                        ***  AGE - Maximum  *** ;
                      else if (sequence in('5')) then
                               do ;
                                   _c7_  = put(_c6_, xxx.) ;
                                   _c11_ = put(_c10_,xxx.) ;
                               end ;
                 end ;
     endcomp ;


     break after prntord / skip ;
run ;

proc print data=outrpt noobs ;
     title 'Output Data Set from PROC REPORT - After COMPUTE' ;
run ;
```

Figure 6

```
                                    Output Data Set from PROC REPORT - After COMPUTE

PRNTORD SEQ LEFTHEAD                                 _C4_ _C5_    _C6_  _C7_       _C8_ _C9_    _C10_ _C11_   FOOLRPT _BREAK_

   1                                                 .             .               .             .               .    PRNTORD
   1    1        N                                   .             .               .             .               .    SEQ
   1    1        N                                118.000 118      .            198.000 198      .               .
   1    2        Mean (S.D.)                         .      0      .               .      0      .               .    SEQ
   1    2        Mean (S.D.)                       52.881  52.9  10.864 (10.86)  54.222  54.2  10.484 (10.48)    .
   1    3        Median                              .             .               .             .               .    SEQ
   1    3        Median                            53.000  53.0    .             54.000  54.0    .               .
   1    4        25th Percentile, 75th Percentile    .             .               .             .               .    SEQ
   1    4        25th Percentile, 75th Percentile  46.000  46.0, 60.000  60.0    47.000  47.0, 61.000  61.0      .
   1    5        Minimum, Maximum                    .             .               .             .               .    SEQ
   1    5        Minimum, Maximum                  28.000  28,   83.000  83      31.000  31,   81.000  81        .
   1                                                 .             .               .             .               .    PRNTORD
   2                                                 .             .               .             .               .    PRNTORD
   2    1        Asian                               .      0      .               .      0      .               .    SEQ
   2    1        Asian                              2       2     1.695   (1.7)   1       1     0.505   (0.5)    .
   2    2        Black                               .      0      .               .      0      .               .    SEQ
   2    2        Black                              5       5     4.237   (4.2)   8       8     4.040   (4.0)    .
   2    3        Caucasian                           .      0      .               .      0      .               .    SEQ
   2    3        Caucasian                        108     108    91.525  (91.5) 180     180    90.090  (90.1)    .
   2    4        Hispanic                            .      0      .               .      0      .               .    SEQ
   2    4        Hispanic                           2       2     1.695   (1.7)   7       7     3.535   (3.5)    .
   2    5        Other                               .      0      .               .      0      .               .    SEQ
   2    5        Other                              1       1     0.847   (0.8)   2       2     1.010   (1.0)    .
   2                                                 .      0      .               .      0      .               .    PRNTORD
```