

Formats in another format

Bettina Ernholt Nielsen, Ferring Pharmaceuticals A/S, Copenhagen, Denmark

ABSTRACT

Have you ever thought about alternative ways of using SAS® formats in your programs? Did you ever use a format to do a calculation or to round your numbers based on the initial value? Often we use formats to read data, display values, group our data or do conversions between numeric and character, but formats can also be used to simplify programs that would have been somewhat complex or lengthy. By carefully constructing our formats to fit our needs we can utilize the powerful SAS format function to accomplish this.

INTRODUCTION

Probably most SAS Programmers use formats in their daily programming when reading and writing data. We also know that SAS provides us with numerous ways to achieve the same result and sometimes a certain programming style results in complex, lengthy, maybe inefficient and hard to read programs. Using formats in your code will help you make simple programs that are more efficient, maintainable and easier to read. This paper will show alternative ways of using the powerful SAS formats to create simple readable programs. The purpose of this paper is to inspire you to use formats throughout your programs and to start thinking about new creative ways of using formats.

In this paper basic knowledge of the FORMAT procedure syntax is assumed, hence no basic syntax will be presented.

EXAMPLE 1: USING SAS FORMATS IN CALCULATIONS

The first example in this paper originated from writing SAS code for calculations of Quality of Life assessments, in particular EQ-5D questionnaire total scores. The EQ-5D is a standardized instrument used as a measure of health related outcome. It provides a simple, generic measure of health for clinical and economic appraisal.

EQ-5D QUESTIONNAIRES:

Subjects in the trial are presented with a questionnaire about quality of life. The questionnaire is divided into subsections about different parts of daily life and the subjects are asked to score each subsection. Subsequently an EQ-5D summary score is calculated based on the coefficients in the table below.

Response option	Mobility(1)	Actual Response	Coefficient
1	I have no problems in walking about		
2	I have some problems in walking about	X	0.069
3	I am confined to bed		0.314
	Self-Care(2)		
1	I have no problems have no problems with self-care		
2	I have some problems washing or dressing my self		0.036
3	I am unable to wash or dress my self	X	0.094
	Usual activities(3) (e.g. work, study, housework, family or leisure activities)		
1	I have no problems with performing my usual activities	X	
2	I have some problems with performing my usual activities		0.123
3	I am unable to perform my usual activities		0.386

Table 1. Sample EQ-5D questionnaire.

Table 1 shows the different subsections and the corresponding coefficients used to calculate the EQ-5D summary score. (Dolan et al, Discussion paper 138)¹

Calculation of EQ-5D summary score:

The EQ-5D summary score is the sum of the coefficients corresponding to the response for each subsection:

$$\text{EQ-5D summary score} = 1 - (\text{Coefficient}_{\text{subsection1}} + \text{Coefficient}_{\text{subsection2}} + \text{Coefficient}_{\text{subsection3}})$$

In the above example the summary score translates to:

$$\text{EQ-5D summary score} = 1 - (0.069 + 0.094 + 0) = 0.837$$

Display 1 shows structure of the data originally collected in the SDTM domain QC.

	SUBJID	VISITNUM	PARAMCD	ORRES
1	0001	1	EQ5D1	2
2	0001	1	EQ5D2	3
3	0001	1	EQ5D3	1
4	0001	4	EQ5D1	1
5	0001	4	EQ5D2	2
6	0001	4	EQ5D3	2

Display 1. Sample data

It seemed cumbersome to write almost endless IF – THEN – ELSE logic (remember there are more subsections in real life) and the final code would end up being long, inefficient and hard to read. The trick here is to see that you have two character values - the question id (subsection) and the original result defining the numeric coefficient. First part (EQ5Dn) is the question identifier, and in this case the same as variable PARAMCD, and the second part (1, 2, 3) is the different response options from the questionnaire.

The following code show how a format constructed and used to compute the total score in a short and effective manner:

```
proc format lib = WORK;
  invalue eq5d
    "EQ5D12" = 0.069 /* Mobility,           Subsection 1 response 2 */
    "EQ5D13" = 0.314 /* Mobility,           Subsection 1 response 3 */
    "EQ5D22" = 0.036 /* Self care,         Subsection 2 response 2 */
    "EQ5D23" = 0.094 /* Self care,         Subsection 2 response 3 */
    "EQ5D32" = 0.123 /* Usual activities, Subsection 3 response 2 */
    "EQ5D33" = 0.386 /* Usual activities, Subsection 3 response 3 */
    other    = 0; /* Add zero for all 1 responses or missing */
run;

proc sort data = EQ5D;
  by SUBJID VISITNUM;
run;

data EQ5DSUM(drop = ORRES);
  set EQ5D;
  by SUBJID VISITNUM;
  retain AVAL; /*-- Save AVAL across iterations of the data step --*/
  if first.VISITNUM then AVAL = .;
  /*-- Create variable that holds question id AND response --*/
  QR = cats(PARAMCD, ORRES);
  /*-- Translate QR into the corresponding coefficient and summarize it --*/
  if not last.VISITNUM then AVAL = sum(AVAL, input(QR, eq5d.));
  else do; /*-- Output summary score by subject and visit number--*/
    AVAL = 1 - sum(AVAL, input(QR, eq5d.));
    PARAMCD = "EQ5DS";
    output;
  end;
run;
```

In order to calculate the EQ-5D summary score all that was done was retaining AVAL and use the CATS function and the SUM function. Now the program is simple to read and easy to maintain if required.

Display 2 shows the resulting data set.

	SUBJID	VISITNUM	PARAMCD	AVAL
1	0001	1	EQ5DS	0.837
2	0001	4	EQ5DS	0.841

Display 2. Calculated data

EXAMPLE 2: ROUNDING WITH SAS FORMATS OR PLEASING YOUR CLINICIANS

This example originated from lengthy discussions with clinicians about the meaning and need for presenting three significant digits in summary tables. The clinicians wanted for example summary statistics on laboratory parameters displayed with 3 significant digits.

THREE SIGNIFICANT DIGITS:

Table 2 shows examples of how three significant digits should be presented. The challenge here is to present numbers for the same variable (column) with different number of decimals depending on the actual value.

Value	Value presented with three significant digits
0.001234	0.00123
0.01234	0.0123
0.1234	0.123
1.234	1.23
12.34	12.3
123.4	123
1234	1230

Table 2. Three significant digits.

Round numbers to three significant digits:

The solution to this problem is beautiful in all its simplicity:

```
AVAL = input(put(AVAL, E9.), E9.);
```

The PUT function returns a value in the format X.XX E YY and when inputting this value you'll get the result in three significant digits.

Pleasing your clinicians:

So the problems were solved with the little magic piece of code, but now it became clear that it wasn't exactly three significant digits the clinicians wanted. They wanted thousands to be presented as is e.g. 1234 as 1234 and not 1230 as the PUT/INPUT functions will return. Hence we modified our code slightly to handle this:

```
AVAL = (AVAL ge 1000) * AVAL + (AVAL lt 1000) * input(put(AVAL, E9.), E9.);
```

Note that you can just use other formats than the E9 if you need a different number of significant digits. This piece of code solved the problem and our clinicians were satisfied. For easy use in DATA steps and SAS® SQL this piece of code can be build into a macro or a function.

CONCLUSION

This paper shows that using formats for less obvious programming tasks and being creative can help to simplify the program code. I hope this paper inspires you to try new techniques as often as you can. It'll broaden your SAS tool box and you'll become a better and more innovative programmer.

REFERENCES

1 Dolan P, Gudex C, Kind P, Williams A. 1995. A Social Tariff for EuroQol: Results from a UK General Population Survey. The University of York, Centre for health economics, York health economics consortium, NHS centre for reviews & dissemination. Discussion paper 138.

ACKNOWLEDGMENTS

Gitte Dam Jepsen, Ferring Pharmaceuticals A/S, Copenhagen

SAS-L

CONTACT INFORMATION

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

Name: Bettina Ernholt Nielsen
Enterprise: Ferring Pharmaceuticals A/S
Address: Kay Fiskers Plads 11
City, State ZIP: 2300 Copenhagen S, Denmark
Work Phone: +45 2878 7419
E-mail: ben@ferring.com
Web: www.ferring.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.