

Determining and Reporting Significant Decimal Places for Continuous Data in Mixed Form

Gary E. Moore, Moore Computing Services, Inc., Little Rock, Arkansas

ABSTRACT

Data of mixed precision is common in clinical trials. Examples of this mixed form are Laboratory, ECG, and Vitals data, where a parameter or type variable defines different results. These results are often recorded to significant decimals appropriate to the data type as determined by the site collecting the data. But, the results are often reported in a common decimal place that may be inappropriate across all of the data types.

This paper presents a simple approach for determining significant decimal places for continuous data collected in mixed form and reporting the appropriate significant decimal places across all data types.

INTRODUCTION

There are common data structures that contain data in mixed form usually defined by a parameter or type variable. The most common data structures of this type are laboratory, electrocardiogram, and vitals. The results defined by a parameter or type variable for these data structures are often collected in varying precisions. For the purposes of this paper, we will deal with lab tests.

In table 1, we have 2 different types of labs, Chemistry and Hematology. Within those lab types we have different lab tests. These lab tests have varying significant digits (total number of digits to the left and right of the decimal). For report programming, we relate significant digits in terms of variable length and significant decimal places.

Labtype	Labtest	Slrslt	Slunit
CHEMISTRY	Albumin	3.6	(g/dL)
CHEMISTRY	Albumin	3.7	(g/dL)
CHEMISTRY	Albumin	3.6	(g/dL)
CHEMISTRY	Alkaline Phosphatase	103	(U/L)
CHEMISTRY	Alkaline Phosphatase	110	(U/L)
CHEMISTRY	Alkaline Phosphatase	105	(U/L)
HEMATOLOGY	Hematocrit	40	(%)
HEMATOLOGY	Hematocrit	40	(%)
HEMATOLOGY	Hematocrit	42	(%)
HEMATOLOGY	Hemoglobin	13.6	(g/dL)
HEMATOLOGY	Hemoglobin	13.3	(g/dL)
HEMATOLOGY	Hemoglobin	12.9	(g/dL)
HEMATOLOGY	WBC	8.69	(x10 ³ /uL)
HEMATOLOGY	WBC	7.08	(x10 ³ /uL)
HEMATOLOGY	WBC	7.43	(x10 ³ /uL)

Table 1. Laboratory Data Example

DESIRED REPORTING OF DECIMAL PLACES

When reporting the original values, it is desired to report the values using the significant decimal places originally recorded in the data.

When reporting summarized values, it is desired to report the results based on the significant decimal places originally recorded in the data. Results that relate directly to the original value such as the Low and High values are reported in the original significant decimal places. Values that are calculated directly from the original value such as Mean and Median are reported in the original significant decimal places + 1 more decimal place. Values that are calculated from calculated values such as standard deviation and standard error are normally reported in the original significant decimal places + 2 more decimal places.

Determining and Reporting Significant Decimal Places for Continuous Data in Mixed Form, continued

Listings	
Results	Recorded Decimal Places
Summaries	
Range (Low Value, High Value)	Recorded Decimal Places
Mean, Median	Recorded Decimal Places + 1
Standard Deviation, Standard Error	Recorded Decimal Places + 2

Table 2. Output Decimal Places

THE PROBLEM

The default SAS formatting will report all lab types and tests using the same formatting within a column.

Labtype	Labtest	SIunit	mean	stderr	median	std	min	max
CHEMISTRY	Albumin	(g/dL)	3.633	0.03333	3.60	0.05774	3.60	3.70
CHEMISTRY	Alkaline Phosphatase	(U/L)	106.000	2.08167	105.00	3.60555	103.00	110.00
HEMATOLOGY	Hematocrit	(%)	40.667	0.66667	40.00	1.15470	40.00	42.00
HEMATOLOGY	Hemoglobin	(g/dL)	13.267	0.20276	13.30	0.35119	12.90	13.60
HEMATOLOGY	WBC	(x10 ³ /uL)	7.733	0.48889	7.43	0.84678	7.08	8.69

Output 1. Default SAS Formatting

We need to be able to control the way the output is formatted within a column, so the rows defined by the lab tests can be reported with different significant decimal places.

Labtype	Labtest	SIunit	mean	stderr	median	std	min	max
CHEMISTRY	Albumin	(g/dL)	3.63	0.033	3.60	0.058	3.6	3.7
CHEMISTRY	Alkaline Phosphatase	(U/L)	106.0	2.08	105.0	3.61	103	110
HEMATOLOGY	Hematocrit	(%)	40.7	0.67	40.0	1.15	40	42
HEMATOLOGY	Hemoglobin	(g/dL)	13.27	0.203	13.30	0.351	12.9	13.6
HEMATOLOGY	WBC	(x10 ³ /uL)	7.733	0.4889	7.430	0.8468	7.08	8.69

Output 2. Desired Formatting

THE SOLUTION

The %SIGDEC macro presented in the Appendix will determine the maximum length and maximum number of decimal places for each lab test. Often, lab data will contain character and numeric variables for the results. It is easier to determine the length and decimal places using the character variable. If the data only contains the numeric version of the result, we should convert it to character and then determine length and decimal places. We use the length and decimal places for each lab test to create a format variable. For reporting, it is easier to control the formatting for the different lab tests using character variables rather than numeric. Using the format variable, we will create new formatted character variables for the lab tests and output these. Calling %SIGDEC will add the new format variables to the original data set.

Labtype	Labtest	SIrslt	SIunit	nSIrslt	sigdec	siglen	sigfmt	sigfmt1	sigfmt2
CHEMISTRY	Albumin	3.6	(g/dL)	3.60	1	3	3.1	4.2	5.3
CHEMISTRY	Albumin	3.7	(g/dL)	3.70	1	3	3.1	4.2	5.3
CHEMISTRY	Albumin	3.6	(g/dL)	3.60	1	3	3.1	4.2	5.3
CHEMISTRY	Alkaline Phosphatase	103	(U/L)	103.00	0	3	3.0	5.1	6.2
CHEMISTRY	Alkaline Phosphatase	110	(U/L)	110.00	0	3	3.0	5.1	6.2
CHEMISTRY	Alkaline Phosphatase	105	(U/L)	105.00	0	3	3.0	5.1	6.2
HEMATOLOGY	Hematocrit	40	(%)	40.00	0	2	2.0	4.1	5.2
HEMATOLOGY	Hematocrit	40	(%)	40.00	0	2	2.0	4.1	5.2
HEMATOLOGY	Hematocrit	42	(%)	42.00	0	2	2.0	4.1	5.2
HEMATOLOGY	Hemoglobin	13.6	(g/dL)	13.60	1	4	4.1	5.2	6.3
HEMATOLOGY	Hemoglobin	13.3	(g/dL)	13.30	1	4	4.1	5.2	6.3
HEMATOLOGY	Hemoglobin	12.9	(g/dL)	12.90	1	4	4.1	5.2	6.3
HEMATOLOGY	WBC	8.69	(x10 ³ /uL)	8.69	2	4	4.2	5.3	6.4
HEMATOLOGY	WBC	7.08	(x10 ³ /uL)	7.08	2	4	4.2	5.3	6.4
HEMATOLOGY	WBC	7.43	(x10 ³ /uL)	7.43	2	4	4.2	5.3	6.4

Output 3. New Data Set

Determining and Reporting Significant Decimal Places for Continuous Data in Mixed Form, continued

We have now created the format variable and additionally the format variables for the +1 and +2 decimal places and can summarize the data.

```
proc summary nway missing data=test;
  class labtype labtest siunit sig;;
  var nsirslt;
  output out=result(drop=_)
    mean=mean stderr=stderr median=median
    std=std min=min max=max
  ;
```

After summarization, we create the character variables for reporting. In this example, we are using the putn function and the format variables to create the new character variables for outputting.

```
data result;
  set result;

  length cmean cstderr cmedian cstd cmin cmax $12;

  cmean=left(putn(mean, sigfmt1));
  cstderr=left(putn(stderr, sigfmt2));
  cmedian=left(putn(median, sigfmt1));
  cstd=left(putn(std, sigfmt2));
  cmin=left(putn(min, sigfmt));
  cmax=left(putn(max, sigfmt));
```

The newly formatted character results after they have been transposed and output with proc report.

Type	Test	Summary	
CHEMISTRY	Albumin (g/dL)	Mean	3.63
		SE	0.033
		Median	3.60
		Std Dev	0.058
		Min	3.6
		Max	3.7
	Alkaline Phosphatase (U/L)	Mean	106.0
		SE	2.08
		Median	105.0
		Std Dev	3.61
		Min	103
		Max	110
HEMATOLOGY	Hematocrit (%)	Mean	40.7
		SE	0.67
		Median	40.0
		Std Dev	1.15
		Min	40
		Max	42
	Hemoglobin (g/dL)	Mean	13.27
		SE	0.203
		Median	13.30
		Std Dev	0.351
		Min	12.9
		Max	13.6
	WBC (x10 ³ /uL)	Mean	7.733
		SE	0.4889
		Median	7.430
		Std Dev	0.8468
		Min	7.08
		Max	8.69

Output 4. New Formatted Output

CONCLUSION

Data should be reported based on the precision that it is recorded. It is relatively easy to accomplish this even when the data structures are in mixed form. Using character values makes easy work of determining length and decimal places for numeric variables. Outputting formatted character values can make reporting result on data of mixed form easier.

ACKNOWLEDGMENTS

I want to thank Michelle Moore for her peer review; John King for making me write a paper; Richard Allen for making me come to PharmaSUG.

CONTACT INFORMATION

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

Name:	Gary E. Moore
Enterprise:	Moore Computing Services, Inc.
Address:	2006 Beckenham Cove
City, State ZIP:	Little Rock, AR 72212
Work Phone:	501-225-8689
E-mail:	gary@moorecsi.net
Web:	www.moorecsi.net

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.

APPENDIX

```

/*****
  %SIGDEC Macro creates format variables for data of varying significant
  decimal places.
  |
  | Created By      :    Gary Moore
  |                  Moore Computing Services, Inc.
  |                  2006 Beckenham Cove
  |                  Little Rock, AR 72212
  |                  501-225-8689
  |                  www.moorecsi.net
  |
  | Name           :    sigdec.sas
  |
  | Parameters     :
  |                 datain    = input dataset
  |                 var       = variable to determine significant decimal places for
  |                 classvars = grouping variables (ie. labtype labtest)
  |                 maxfmtlen = maximum format length for converting numeric to
  |                           character
  |                 fmtname   = output variable containing complete format
  |                 sigdecvar = output var containing significant decimal places
  |                 siglenvar = output var containing format length
  |
  *****/
%macro sigdec(datain=,
              var=,
              classvars=,
              fmtname=sigfmt,
              maxfmtlen=12,
              sigdecvar=sigdec,
              siglenvar=siglen);

data &datain;
set &datain;

/*****
dummy merging variable in case
no class variables are defined
*****/
mergeby_ = 1;

proc summary nway missing data=&datain;
class mergeby_ &classvars &var;
output out=sigvars(drop=_:);
run;

/*****
check to see if variable is
character or numeric
*****/

%let dsid=%sysfunc(open(sigvars,i));
%let varnum=%sysfunc(varnum(&dsid,&var));
%let vartyp=%sysfunc(vartype(&dsid,&varnum));
%let rc=%sysfunc(close(&dsid));

%put Note: Variable &varnum is &vartyp;

```

Determining and Reporting Significant Decimal Places for Continuous Data in Mixed Form, continued

```

data sigvars;
  set sigvars;

  /*****
  convert to character if numeric
  *****/

  %if &vartyp = N %then
    %do;
      sigvar = trim(left(put(&var,best&maxfmtlen..)));
    %end;
  %else
    %do;
      sigvar = &var;
    %end;

  /* Length */
  &siglenvar = length(compress(sigvar));

  /* Decimals */
  if index(sigvar, '.')
    then &sigdecvar = length(sigvar) - index(sigvar, '.');
    else &sigdecvar = 0;

  /* Maximums */
  proc summary nway missing data=sigvars;
    class mergeby_ &classvars;
    var &sigdecvar &siglenvar;
    output out=sigvar(drop=_) max=&sigdecvar &siglenvar;

  proc sort data=&datain;
    by mergeby_ &classvars;

  /* Merge back to original data */
  data &datain(drop=mergeby_);
    merge &datain
      sigvar;

  by mergeby_ &classvars;

  /* create format variable */
  &fmtname = compress(put(&siglenvar,&maxfmtlen..)
    ||'.'||put(&sigdecvar,&maxfmtlen..));

  /* create format variable with decimal + 1 */
  if &sigdecvar = 0
    then &fmtname.1 = compress(put(&siglenvar+2,&maxfmtlen..)
      ||'.'||put(&sigdecvar+1,&maxfmtlen..));
    else &fmtname.1 = compress(put(&siglenvar+1,&maxfmtlen..)
      ||'.'||put(&sigdecvar+1,&maxfmtlen..));

  /* create format variable with decimal + 2 */
  if &sigdecvar = 0
    then &fmtname.2 = compress(put(&siglenvar+3,&maxfmtlen..)
      ||'.'||put(&sigdecvar+2,&maxfmtlen..));
    else &fmtname.2 = compress(put(&siglenvar+2,&maxfmtlen..)
      ||'.'||put(&sigdecvar+2,&maxfmtlen..));

  run;

  %mend sigdec;

```