We use INPUT function to convert a character variable to numeric. Problems, however, can occur when the variable contains values that cannot be converted to numeric (like '20+', '<30', '100-200') by INPUT. We all see the unwelcome message of "Invalid numeric data" too many times.

VB has a Boolean function, IsNumeric(), which takes a character variable and returns true or false based on whether the character variable can be converted to numeric. SAS has a similar macro function, %DATATYP, to check the data type of a macro variable. Interestingly, there is no counterpart for data step processing.

Presented here is a data step macro function, %isnum, mimicking the IsNumeric() in VB. Which can be used in if logic to check the data type before calling INPUT function. Like

```
if %isnum(charval) then nval=input(charval,best);
```

And we say good-bye to "Invalid numeric data", forever.

To subset a data set into numeric and character, use

```
data nval cval;
  set source;
  if %isnum(charval) then output nval;
  else output cval;
run;
```

Advantage: The macro is nothing but a logic evaluation. The macro call is simple and natural to human thinking. And we choose the name %isnum to enhance readability.

Efficiency issues: we divide all possible conditions in parallel "or" logic clauses ordered by their frequency of occurrences. Same rule also applied to the sub-logics within the "or" clauses. So the macro runs very fast for real data, even though the code look complicated.

Note: Do not use for scientific notation, like "1.5E-9". Which is valid argument for INPUT but %IsNum will treated it as character. We did not try to handle scientific notation, as it is rarely used in clinical trial data. And obviously, the logic is too complicated to fit in an if statement.

Your questions and comments are very much appreciated. And we can be reached at:

Fuping Peng
Eisai Medical Research, Inc
55 Challenger Rd
Ridgefield Park, NJ 07660
201-287-2368
fuping_peng@eisai.com
%macro isnum(str);
    verify(trim(left(&str)),'0123456789')=0 or /*number only*/
    verify(trim(left(&str)),'0123456789.')=0
    and not indexc(substr(&str,indexc(&str,'.')+1), '.') or /*allow only one "."*/
    verify(trim(left(&str)),'0123456789.+-')=0
    and not indexc(substr(&str,indexc(&str,'.')+1), '.')
    and (indexc(&str,'+-')=1
    and not indexc(substr(&str,2),'+-')               /*allow only one leading '+' or '-'*/
    and indexc(&str,'0123456789.') > 1 ) or      /* '+-' must followed by number*/
    compress(&str='')                                /*'', ' ', or multiple ' ' is numeric*/
%mend;

/*sample call
data c;
    length str $30;
    str='.';                 output;
    str='';                         output;
    str='+';                         output;
    str='+.';                         output;
    str='+.1';                        output;
    str='+.0';                        output;
    str='  ';                output;
    str='123456789';        output;
    str='  1234567899  ';   output;
    str='1234 56789';       output;
    str='1234,56789';       output;
    str='+123456789';       output;
    str='+-123456789';      output;
    str='+12345-6789';      output;
    str='12345+6789' ;      output;
    str='+123.';          output;
    str='.12.323';          output;
    str='123.23.';          output;
    str='-123.23';          output;
    str='-12323';          output;
    str='-12323-';          output;
    str='12323-';           output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='.123.';            output;
    str='0000.123400';      output;
    str='assdfas400';      output;
    str='1.5E-9';           output;
    run;

data n;
length num 8;
set c;
if %isnum(str) then do;
   num=input( str, best.);
   put str 'is numeric';
end;
else  put str 'is character';
if %isnum(str);
run;
*/
SAS Log:
3107  data n;
3108   length num 8;
3109   set c;
3110   if %isnum(str) then do;
3111      num=input( str, best.);
3112      put str 'is numeric';
3113    end;
3114   else put str 'is character';
3115   if %isnum(str);
3116  run;

. is numeric
   is numeric
+ is character
+. is numeric
+.1 is numeric
+.0 is numeric
   is numeric
123456789 is numeric
1234567899 is numeric
1234 56789 is character
1234,56789 is character
+123456789 is numeric
+-123456789 is character
+12345-6789 is character
12345+6789 is character
123.23+ is character
-123.23 is numeric
-12323 is numeric
-12323- is character
12323- is character
.12.323 is character
123.23.45.67 is character
00.123 is numeric
.123 is numeric
+.123 is numeric
.123. is character
.123+ is character
+123. is numeric
0000.123400 is numeric
assdfsad400 is character
1.5E-9 is character
NOTE: There were 31 observations read from the data set WORK.C.
NOTE: The data set WORK.N has 16 observations and 2 variables.