



New Argument for Use with the COMPRESS function.

Rosalind Gusinow

You might be familiar with the COMPRESS function. It has been around for quite a while. It allows you to remove selected characters from a character value. If you execute this function with only one argument (a character value), the default action is to remove blanks from the string. By specifying a second argument, you can specify which characters you want to remove from the string. Here is an example of how the previous (SAS 8) version of this function is used to remove all uppercase and lowercase letters, periods, dashes, and blanks from a character value:

```
data numbers_only;
  input ID $10.;
  number_part = compress(lowercase(ID), 'abcdefghijklmnopqrstuvwxyz.- ');

datalines;
12 ab c.-34
abc1 2 - .89
;
```

The SAS 9.1 upgrade allows a third argument that lets you specify character classes (i.e., letters, digits, punctuation, white space, etc.) to remove from the character value. In addition, other modifiers change the way this function works. You can use one modifier to tell the function to remove all characters except the ones specified and another modifier to trim trailing blanks or ignore cases. Following are a few of the more useful modifiers:

Selected list of COMPRESS modifiers (uppercase or lowercase)

a	Adds uppercase and lowercase letters
d	Adds numerals (digits)
i	Ignores case
k	Keeps listed characters instead of removing them
s	Adds space (blank, tabs, lf, cr) to the list
p	Adds punctuation

```

Options nocenter;

DATA numbers_only;
  input ID $10.;
  ID_length=length(ID);

  /* tradition use of the Compress function to eliminate letters and punctuation*/
  number_part = compress(lowercase(ID), 'abcdefghijklmnopqrstuvwxyz.- ');
  np_length=length(number_part);

  /* third argument: l-lower case,
                     u-uppercase */
  number_part_no_upper_or_lower = compress(ID,, 'lu');
  npnuol_length=length(number_part_no_upper_or_lower);

  /* third argument: a-alphabet(upper and lower),
                     s-adds space characters to the list of characters */
  number_part_no_alpha_or_spaces = compress(ID,, 'as');
  npnaos_length=length(number_part_no_alpha_or_spaces);

  /* third argument: p-punctuation */
  number_part_no_punctuation = compress(ID,, 'p');
  np_length=length(number_part_no_punctuation);

  number_part_only = compress(ID,, 'asp');
  npo_length=length(number_part_only);

  retain crlf_var '0d'x;

  number_part_crlf = compress(ID,crlf_var);
  npcrlf_length=length(number_part_crlf);

  number_part_crlf_only = compress(ID,'0d'x,'k');
  npcrlfo_length=length(number_part_crlf_only);

datalines;
12 ab Bc.-34
abc1 2 B- .89
;
Run;

PROC PRINT data=numbers_only;
  Title 'numbers_only';
RUN;

```