

DATA Step

Why shouldn't I use TRANWRD or TRANSLATE to remove characters from a string?

The TRANWRD function can be used to replace a 'word' within a character string, and the TRANSLATE function can be used to replace any of a list of characters. However, the minimum length of a character string is 1, therefore it is not possible to remove a word or character by replacing it with a 'null' string.

Consider the following data:

```
data cleansing ;
  length str $ 6 ;
  str = 'B+' ;
  output ;
  str = 'B/C' ;
  output ;
  str = '-A' ;
  output ;
run ;
```

Attempting to remove the special characters + - / using the TRANWRD function with an apparent 'null' string iteratively:

```
data cleaned1 ;
  set cleansing ;
  str = tranwrd(str, '+', '') ;
  str = tranwrd(str, '/', '') ;
  str = tranwrd(str, '-', '') ;
  len = length(str) ;
  put '-->' str '<-- Length: ' len ;
run ;
```

Examining the LOG:

```
-->B <-- Length: 1
-->B C <-- Length: 3
-->A <-- Length: 2
```

The embedded and leading blanks are still included in the length although only the embedded blank is clearly evident; the trailing blank is masked by the remaining 'padding' blanks in the

DATA Step

fixed-length string. The leading blank is masked by the left-alignment of the 'standard' format.

Attempting the exercise with the TRANSLATE function (character by character replacement), has the same effect:

```
data cleaned2 ;
  set cleansing ;
  str = translate(str, ' ', '+/-') ;
  len = length(str) ;
  put '-->' str '<-- Length: ' len ;
run ;
```

Examining the LOG:

```
-->B <-- Length: 1
-->B C <-- Length: 3
-->A <-- Length: 2
```

To remove, rather than replace, the characters, use the COMPRESS function:

```
data properly_cleaned ;
  set cleansing ;
  str = compress(str, '+/-') ;
  len = length(str) ;
  put '-->' str '<-- Length: ' len ;
run ;
```

Examining the LOG:

```
-->B <-- Length: 1
-->BC <-- Length: 2
-->A <-- Length: 1
```

Unique solution ID: #1014

Author: Alan D Rudland

Last update: 2017-05-12 08:48