

DATA Step

Can I stop SAS from reading lines from a csv file where all values are missing?

Sometimes when generating a CSV file, e.g. from an Excel Spreadsheet there may be lines in the file which consist of nothing but the field separators, e.g.:

```
','','',''
```

Using normal INFILE / INPUT statements these rows are still included in the dataset, however all values are missing.

To remove these spurious lines, it is possible to check that there is no valid data, by first removing the field separators, then determining the length of the string.

The _INFILE_ keyword can be used to access the Input Buffer where each line of raw data is stored as it is being processed.

The COMPRESS keyword can be used to remove specified characters from a character string, e.g.: compress(_infile_,',') to remove all commas.

The LENGTH function will return the position of the last non-blank character in a character string, however, if the character string is 'missing' then the function returns a length of 1. To return a zero length for 'missing' strings use the LENGTHN function instead.

Submitting the following:

```
data no_empty_lines ;
  infile datalines dsd ;
  input name $
         id
         gender $
         ;
  if lengthn(compress(_infile_',')) = 0 then delete ;
  datalines ;
Alan,1,M
Joe,2,M
Jo,3,F
, ,
, ,
;
run ;
```

Generates a dataset with the desired three observations.

Unique solution ID: #1026

DATA Step

Author: Alan D Rudland

Last update: 2017-05-10 11:57