

DATA Step

How does the 'SOUNDS LIKE' (=*) operator work in a WHERE statement?

The SOUNDS LIKE (=*) operator uses a soundex algorithm to simplify a word for comparison purposes. The algorithm is predicated on English however, and is less accurate for other languages.

Firstly retain the first letter and discard the following letters: A E H I O U W Y

Secondly, substitute the numbers as follows to the letter groupings:

1. B F P V
2. C G J K Q S X Z
3. D T
4. L
5. M N
6. R

finally, if two or more adjacent letters (prior to discarding letters) have the same number classification resulting from the second step, then de-duplicate.

The same logic is used by the SOUNDSEX function, e.g.:

```
data al ;  
  length name name_sound $ 10 ;  
  name = 'Alan' ;  
  name_sound = soundex(name) ;  
  output ;  
  name = 'alun' ;  
  name_sound = soundex(name) ;  
  output ;  
  name = 'ALLAN' ;  
  name_sound = soundex(name) ;  
  output ;  
  name = 'Allen' ;  
  name_sound = soundex(name) ;  
  output ;  
  name = 'Alyn' ;  
  name_sound = soundex(name) ;  
  output ;  
  name = 'Alwyn' ;  
  name_sound = soundex(name) ;  
  output ;  
  name = 'David' ;  
  name_sound = soundex(name) ;  
  output ;  
run ;
```

```
data soundslike ;
```

DATA Step

```
set al ;  
  where name =* 'ALAN' ;  
run ;
```

Note that both the SOUNDSEX function and the SOUNDS LIKE (=*) operator are case-insensitive!

Unique solution ID: #1007

Author: Alan D Rudland

Last update: 2017-02-23 14:56