

Use a Format in a Subsetting IF

William Kreuter, University of Washington, Seattle, WA 98195-8853

The base language's IN operator was a welcome addition a few years ago, but it lacks versatility because individual elements must be separately specified. Thus, a logical comparison formed from an inclusive range or ranges of values plus some individual values can still be cumbersome to write.

But there's always been a practical alternative which makes this easy. Just put all the ranges and values you desire into a value format, and then establish the format such that it can only return a "1" or a "0". The formatted string itself can then be applied directly in an IF-THEN or a subsetting IF by use of the PUT function.

EXAMPLE.

Suppose you need to subset on the following values of X: 1 to 3, 9, 11, 15 to 19, 300 to 1000. Use the following.

```
proc format;
  value subset
    1-3,9,11,15-19,300-1000='1'
    other='0';
data datastep;
/* ... */
  if put(x,subset.);
```

The PUT function returns the string "1", if X has a value within the list to be subsetted, or "0" otherwise. This string is passed as the target of the subsetting IF, which then does an implicit character-to-numeric conversion and appropriately treats the resulting numeric 1 (retain in subset) or 0 (reject from subset).

I propose that this is more straightforward than alternative methods of obtaining the same result, since all values to be subsetted can be specified in a succinct, comma-delimited list.

William Kreuter
University of Washington 358853
Seattle, WA 98195-8853
billyk@u.washington.edu