Subject: Re: MODE in IDL?
Posted by JD Smith on Wed, 25 Jan 2006 17:14:39 GMT
View Forum Message <> Reply to Message

On Tue, 24 Jan 2006 20:04:13 -0700, David Fanning wrote:

> array = [1, 1, 2 , 4, 1, 3, 3, 2, 4, 5, 3, 2, 2, 1, 2, 6]
> h = Histogram(array, MIN=Min(array))
> bigfreq = Max(h)
> mode = Where(h EQ bigfreq) + Min(array)
> Print, mode
> 2

Just a hint on your HISTOGRAM usage... this might be slightly preferred, since it skips the WHERE and MIN:

```
array = [1, 1, 2 , 4, 1, 3, 3, 2, 4, 5, 3, 2, 2, 1, 2, 6]
void=max(histogram(array,OMIN=mn),mxpos)
mode=mn+mxpos
```

This method of course will be *very* problematic if you have, e.g.:

```
array = [1, 1, 2 , 4, 1, 3, 3, 2, 4, 5, 3, 2, 2, 1, 2, 6, 10000000]
```

Another option, if you worry about this, would be to re-cast as a sorting problem, using the method discussed in a recent thread:

```
array=array[sort(array)]
wh=where(array ne shift(array,-1),cnt)
if cnt eq 0 then mode=array[0] else begin
  void=max(wh-[-1,wh],mxpos)
  mode=array[wh[mxpos]]
endelse
print,mode
```

Both methods will give you the lowest number in the case of ties for the mode. The second will be slower, but more robust against large dynamic range in your array. You could use both, deciding which to use by the min/max of the array.

JD