Subject: Re: MODE in IDL?
Posted by David Fanning on Wed, 25 Jan 2006 03:04:13 GMT
View Forum Message <> Reply to Message

Jonathan Greenberg writes:

> I can't seem to find the function to calculate the mode of an array -- how
> do I do this in IDL?

I'm trying to remember back to grade school or somewhere, but
isn't the mode the maximum of the frequency distribution. I
remember that the mode isn't unique, because two numbers
could have the same frequency in the sample (bimodal, I guess)
and it is theoretically possible to have no mode (all frequencies
are the same).

But saying all that, if you have an integer array, I would
think the mode is calculated like this:

```
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
```

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Re: MODE in IDL?
Posted by Jonathan Greenberg on Wed, 25 Jan 2006 03:13:59 GMT
View Forum Message <> Reply to Message

David:
That is the defintion of mode -- it really surprised me this isn't built
into IDL, since its a common and basic enough statistic.

--j

"David Fanning" <davidf@dfanning.com> wrote in message
news:MPG.1e40988f4505ca03989b37@news.frii.com...

> Jonathan Greenberg writes:
>
>> I can't seem to find the function to calculate the mode of an array --
>> how
>> do I do this in IDL?
>
> I'm trying to remember back to grade school or somewhere, but
> isn't the mode the maximum of the frequency distribution. I
> remember that the mode isn't unique, because two numbers
> could have the same frequency in the sample (bimodal, I guess)
> and it is theoretically possible to have no mode (all frequencies
> are the same).
>
> But saying all that, if you have an integer array, I would
> think the mode is calculated like this:
>
> 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
>
> Cheers,
>
> David
>
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming: http://www.dfanning.com/

Subject: Re: MODE in IDL?
Posted by David Fanning on Wed, 25 Jan 2006 03:20:05 GMT
View Forum Message <> Reply to Message

Jonathan Greenberg writes:

> That is the defintion of mode -- it really surprised me this isn't built
> into IDL, since its a common and basic enough statistic.

Maybe it is and I've just never run into it. I relied on
you to do the literature search. :-)

Cheers,

David

--
David Fanning, Ph.D.
Fanning Software Consulting, Inc.
Coyote's Guide to IDL Programming: http://www.dfanning.com/

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