Subject: weighted average
Posted by gunvicsin11 on Tue, 10 Dec 2013 06:41:37 GMT
View Forum Message <> Reply to Message

Hello everyone,
I have data set of 800 data points spreaded in 20 years, I have to take six monthly mean of the data set, but I need to take weighted mean. My preference is that , while taking six monthly mean more weightage has to be given to those values which are more in numbers, less weightage to those values which lesser in numbers. How can I do this, is there any idl routine that does this. Please help me out in this regard.
thanking you in advance
sid

Subject: Re: weighted average
Posted by on Tue, 10 Dec 2013 09:29:40 GMT
View Forum Message <> Reply to Message

Den tisdagen den 10:e december 2013 kl. 07:41:37 UTC+1 skrev sid:
> Hello everyone,
>
> I have data set of 800 data points spreaded in 20 years, I have to take six monthly mean of the data set, but I need to take weighted mean. My preference is that , while taking six monthly mean more weightage has to be given to those values which are more in numbers, less weightage to those values which lesser in numbers. How can I do this, is there any idl routine that does this. Please help me out in this regard.
>
> thanking you in advance
>
> sid

weighted_mean = total(data*weights)/total(weights)

Subject: Re: weighted average
Posted by Craig Markwardt on Tue, 10 Dec 2013 16:15:34 GMT
View Forum Message <> Reply to Message

On Tuesday, December 10, 2013 1:41:37 AM UTC-5, sid wrote:
> Hello everyone,
>
> I have data set of 800 data points spreaded in 20 years, I have to take six monthly mean of the data set, but I need to take weighted mean. My preference is that , while taking six monthly mean more weightage has to be given to those values which are more in numbers, less weightage to those values which lesser in numbers. How can I do this, is there any idl routine that does this. Please help me out in this regard.

Mats has the right idea. But usually the analyst has a rationale for assigning weights. For example, if each data point has an error bar (gaussian), then weights is typically assigned (1/error^2).

If you want to assign weight based on the value ("more in numbers"), you can do that, but why? There is a danger in assigning weights arbitrarily. By emphasizing the larger values, you will bias your mean value toward larger values. That bias may or may not be reasonable. In the extreme limit, why not just take the maximum of all of your data points?
