Subject: Re: non-integer binsize
Posted by David Fanning on Thu, 05 May 2005 00:54:43 GMT
View Forum Message <> Reply to Message

patrick.gatlin@msfc.nasa.gov writes:

> This may be a very elementary question, but I was wondering if the
> histogram function can accept a binsize < 1?

Since the example in the on-line help for HIST_2D uses a
bin size of 0.02, I think you might be in good shape here. ;-)

> I am using the hist_2d function to determine the density field of some
> lat/lon locations in a dataset. These values range 4 deg lat and 4 deg
> lon. In order to determine the # of points per square km, I have setup
> the hist_2d function as follows:
>
> IDL> result=hist_2d(lon,lat,max1=lonmax,min1=lonmin,max2=latmax,$
> min2=latmin,bin1=1/111.,bin2=1/111.)
>
> where the size of bin1 and bin2 are the approximate conversions from km
> to degrees (1km=.09 deg).
> Since hist_2d needs two bins, then is the result equal the number of
> points within a box that is 1/111. long on all sides? Thus am assuming
> result will give me the # of points per square kilometer. Is this a
> correct assumption based upon the way hist_2d works?

This explanation reminds me of editing my middle son's
big essay last night (sigh...), but I think you've probably
got the gist of it, anyway. :-)

I think a better way of looking at it is that the box is
1/111. times the number of bins between the min and max values,
but I think you are correct that the numbers you get back
represent the number of points in a square kilometer.

Cheers,

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

Subject: Re: non-integer binsize
Posted by JD Smith on Thu, 05 May 2005 16:44:47 GMT

On Wed, 04 May 2005 17:28:04 -0700, patrick.gatlin wrote:

> This may be a very elementary question, but I was wondering if the
> histogram function can accept a binsize < 1?
>
> I am using the hist_2d function to determine the density field of some
> lat/lon locations in a dataset. These values range 4 deg lat and 4 deg
> lon. In order to determine the # of points per square km, I have setup
> the hist_2d function as follows:
>
> IDL> result=hist_2d(lon,lat,max1=lonmax,min1=lonmin,max2=latmax,$
> min2=latmin,bin1=1/111.,bin2=1/111.)
>
> where the size of bin1 and bin2 are the approximate conversions from km
> to degrees (1km=.09 deg).
> Since hist_2d needs two bins, then is the result equal the number of
> points within a box that is 1/111. long on all sides? Thus am assuming
> result will give me the # of points per square kilometer. Is this a
> correct assumption based upon the way hist_2d works?

If you peek inside HIST_2D (or HIST_ND for that matter), you'll see that the way non-integer bin-sizes work is by first converting the data to some suitable integer representation over the range of interest, typically where the binsize is normalized to 1, and then taking the histogram. This is also quite likely how HISTOGRAM itself works internally. You might also like to read up on the rounding issues you can have when working with floating point histograms:

http://www.dfanning.com/math_tips/razoredge.html

JD

Subject: Re: non-integer binsize
Posted by patrick.gatlin on Thu, 05 May 2005 20:23:53 GMT

Thank you David and JD for the help and some insight into the histogram. I wanted to make sure I was not misinterpreting the hist_2d output.
At least you didn't have to break out the red pen on this one David.
;-)
Thanks again!
