Subject: histogram bin edges
Posted by Josh Sixsmith on Sun, 23 Dec 2012 02:56:24 GMT
View Forum Message <> Reply to Message

Hi, I'm curious about whether IDL is inclusive or exclusive in regards to the final bin. The help documentation describes that the right hand side of a bin is exclusive, but what about the final bin?

For example, using the following sample data

a = [0.72244781, 0.20885457, 0.38053078, 0.89579923, 0.93703798,$
    1.   , 0.22754776, 0.11365818, 0.38424101, 0.1741128 ,$
    0.63094614, 0.00123615, 0.06025917, 0.78652067, 0.1001857 ,$
    0.80492211, 0.80564817, 0.83369342, 0.94378603, 0.75453023]

ha = histogram(a, nbins=10, binsize=0.1, min=0, reverse_indices=ria, locations=loca, omax=omaxa, omin=omina)
IDL> print, loca
    0.00000   0.10000   0.20000   0.30000   0.40000   0.50000
    0.60000   0.70000   0.80000   0.90000

IDL> print, ha
        2   3   2   2   0   0
        1   3   4   3

This result looks fine and i interpret the locations as

 [0,0.1),[0.1,0.2),[0.2,0.3),[0.3,0.4),[0.4,0.5),[0.5,0.6),[0 .6,0.7),[0.7,0.8),
[0.8,0.9),[0.9,1.0]

indicating the last bin as inclusive as the value of 1.0 is included in the histogram.

However i feel that this might actually be a case of misrepresentation of floating point numbers, as mentioned in the example given in http://www.idlcoyote.com/math_tips/razoredge.html

If array 'a' is double:
ad = double(a)
had = histogram(ad, nbins=10, binsize=0.1d, min=0.0d, reverse_indices=riad, locations=locad, omin=ominad, omax=omaxad)
print, had
        2   3   2   2   0   0
        1   3   4   2

which suggests that the last bin is exclusive and that all bins are exclusive

 [0,0.1),[0.1,0.2),[0.2,0.3),[0.3,0.4),[0.4,0.5),[0.5,0.6),[0 .6,0.7),[0.7,0.8),
[0.8,0.9),[0.9,1.0)

For integer data this also seems to be the case
IDL> print, locbHowever, for integer data this doesn't seem to be the case.
b = indgen(11)
hb = histogram(b, nbins=10, min=0, reverse_indices=rib, locations=locb, omax=omaxb, omin=ominb)
IDL> print, locb
     0   1   2   3   4   5   6   7   8   9
IDL> print, hb
        1   1   1   1   1   1
        1   1   1   1

IDL> print, rib
       11   12   13   14   15   16
       17   18   19   20   21   0
        1   2   3   4   5   6
        7   8   9
The index '10' for the value 10 is not included

So if i use the max keyword, i would assume that this would be an upper limit of values to be included in the histogram. Does this then make the last bin inclusive, or will it create a bin that will not only include the max value but potentially include values slightly higher than the specified max. This might only occur if the binsize is set.

Any clarification would be greatly appreciated.

Cheers
Josh