Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by JD Smith on Thu, 12 Jun 2003 22:25:04 GMT
View Forum Message <> Reply to Message

On Thu, 12 Jun 2003 14:54:13 -0700, Tim Robishaw wrote:

> Thanks a bunch, folks. Your responses were very helpful. Much
> appreciated.
>
> -Tim.
>
> Paul van Delst wrote:
>
>> The result isn't wrong. Your assumptions about the numerical accuracy
>> are. I'm amazed you're surprised at getting a result of
>> 1.9999980926513671875 for the expression (-5.40-(-5.50))/0.05, Even
>> double precision won't help you here:
>>
>> IDL> print, (-5.40d0-(-5.50d0))/0.05d0, format='(f21.19)'
>> 1.9999999999999928946
>
> Paul:
>
> First of all, let me thank you for your amazement at my lack of
> understanding. That's not a very friendly way to encourage people to
> post questions to your group. Secondly, I was not surprised that your
> example wasn't an integer; rather, I discovered this was exactly the
> reason HISTOGRAM *wasn't* working as I expected it to. I was surprised
> that HISTOGRAM was subtracting the MIN and then dividing by the BINSIZE
> when this is bound to goof up values at the hairy edges of bins. My
> question was whether or not there was a way to get around this problem
> should you be expecting it.

Don't sweat it... this friendly gibing tone is part and parcel of the group, and what keeps you coming back for more ;).

The problem can be re-phrased as: "What should histogram do with values exactly on a bin boundary?" Even if the BINSIZE is 1, the problem can occur. For floating-point values, this question is tricky, since the internal process of transforming the data into logical bin numbers can move it from one side of the boundary to another. Luckily, there is no ambiguity when binning integer data types: if an integer falls precisely on the bin boundary, it is put into the higher-valued bin. If you were really concerned about this problem (and I can't think of too many instances were I'd be), you could do the integer conversion yourself, perhaps rounding to the nearest bin boundary for values close (< eps) to such a boundary, and then perform your histogram on the resultant integer data.

Good luck,

JD