
Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [Craig Markwardt](#) on Thu, 12 Jun 2003 07:04:28 GMT
[View Forum Message](#) <> [Reply to Message](#)

timrobishaw@yahoo.com (Tim Robishaw) writes:

```
...
> IDL> print, histogram([-5.50,-5.45,-5.40,-5.35,-5.30,-5.25],min=-5.50,bin size=0.05)
>      1      2      0      2      0      1
>
> Wait a minute, this should be a uniform distribution!
...
>
> Well, there ya go. It's a roundoff error problem that results from
> trying to balance the values on a razor's edge... the subtraction and
> division knock a few values off balance. But, the result is still
> WRONG and I just don't know enough about roundoff error to know if
> this is an insurmountable problem (but my educated guess is: yes.) Is
> there some clever way to make HISTOGRAM behave properly in such
> situations?
```

Partial solution number one: use double precision.

Partial solution number two: multiply all values by (1+EPS)
where EPS = (MACHAR()).EPS (or equivalent double version)
(assumes you always want to round "up" to the next bin)

Partial solution number three: add a random deviate to each value
(doesn't solve the razor's edge problem per se, but reduces the
chances that a human-entered quantity will land on a bin-edge, and
reduces the bias of always rounding "up" to the next bin)

Partial solution number four: work in powers of 2 instead of multiples
of 0.05.

Partial solution number five: learn to live with it.

Good luck!
Craig

--

Craig B. Markwardt, Ph.D. EMAIL: craigmnet@cow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response

Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [R.G. Stockwell](#) on Thu, 12 Jun 2003 11:33:10 GMT
[View Forum Message](#) <> [Reply to Message](#)

```

"Tim Robishaw" <timrobishaw@yahoo.com> wrote in message
news:405594fa.0306112244.6fcd81d6@posting.google.com...
> So I've been reading up on HISTOGRAM and how it is optimized to be
> wicked fast.
> Also, I'm discovering the fancy things one can do with reverse
> indices.
> I've even read JD Smith's exegesis on the topic:
> "HISTOGRAM: The Breathless Horror and Disgust"
>
> So, just when I thought it was safe to start using HISTOGRAM with the
> frequency that Californians use LIKE, I was brought this scary result
> by the guy on the other side of my wall (his name is Tiberius):
>
> IDL> print,
> histogram([-5.50,-5.45,-5.40,-5.35,-5.30,-5.25],min=-5.50,binsize=0.05)
>      1      2      0      2      0      1
>
> Wait a minute, this should be a uniform distribution!

```

No, histogram gives the correct result. Check out what you put into the histogram function.

```
IDL> print, [-5.50,-5.45,-5.40,-5.35,-5.30,-5.25],format='(f50.25)'  
-5.500000000000000000000000000000  
-5.4499998092651367000000000000  
-5.4000000953674316000000000000  
-5.3499999046325684000000000000  
-5.3000001907348633000000000000  
-5.2500000000000000000000000000
```

So, if such razor edge stuff is a concern of yours, preprocess the data and use integers (i.e. round to integers). This also applies to any conditional tests of a float (for instance, for $i = 0.1, 10, 0.001$ do begin... etc).

```
IDL> print, histogram(
round([-5.50,-5.45,-5.40,-5.35,-5.30,-5.25]/0.05),min=-5.50/ 0.05,binsize=1)
1 1 1 1 1 1
```

Cheers,
bob

Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [James Kuyper](#) on Thu, 12 Jun 2003 13:53:58 GMT
[View Forum Message](#) <> [Reply to Message](#)

Tim Robishaw wrote:

...
> So, just when I thought it was safe to start using HISTOGRAM with the
> frequency that Californians use LIKE, I was brought this scary result
> by the guy on the other side of my wall (his name is Tiberius):
>
> IDL> print, histogram([-5.50,-5.45,-5.40,-5.35,-5.30,-5.25],min=-5.50,bin size=0.05)
> 1 2 0 2 0 1
>
> Wait a minute, this should be a uniform distribution!
>
> Now, I admit that Tiberius contrived this example with the intent to
> cause harm to HISTOGRAM by placing the values at the exact boundaries
> of each bin. But, we didn't EXPECT it to break! Honest. After

If you expect to get a lot of values that are within round-off error of being exactly on the boundary values, then your boundary values have been poorly placed. You should use min=-5.525 for this case.

Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [David Fanning](#) on Thu, 12 Jun 2003 14:30:40 GMT
[View Forum Message](#) <> [Reply to Message](#)

Craig Markwardt writes:

> Partial solution number one: use double precision.
>
> Partial solution number two: multiply all values by (1+EPS)
> where EPS = (MACHAR()).EPS (or equivalent double version)
> (assumes you always want to round "up" to the next bin)
>
> Partial solution number three: add a random deviate to each value
> (doesn't solve the razor's edge problem per se, but reduces the
> chances that a human-entered quantity will land on a bin-edge, and
> reduces the bias of always rounding "up" to the next bin)
>
>
> Partial solution number four: work in powers of 2 instead of multiples
> of 0.05.
>
> Partial solution number five: learn to live with it.

It's been a few months. Everyone ought to have another

go at the Sky is Falling article and the good links therein:

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

Cheers,

David

--

David W. Fanning, Ph.D.

Fanning Software Consulting, Inc.

Phone: 970-221-0438, E-mail: david@dfanning.com

Coyote's Guide to IDL Programming: <http://www.dfanning.com/>

Toll-Free IDL Book Orders: 1-888-461-0155

Subject: Re: HISTOGRAM and the Razor's Edge.

Posted by [Paul Van Delst\[1\]](#) on Thu, 12 Jun 2003 14:40:50 GMT

[View Forum Message](#) <> [Reply to Message](#)

Tim Robishaw wrote:

```
>
> IDL> print, ([-5.50,-5.45,-5.40,-5.35,-5.30,-5.25]-(-5.50))/0.05,
> format='(f21.19)'
> 0.00000000000000000000
> 1.0000038146972656250
> 1.9999980926513671875
> 3.0000019073486328125
> 3.9999961853027343750
> 5.00000000000000000000
>
> Well, there ya go. It's a roundoff error problem that results from
> trying to balance the values on a razor's edge... the subtraction and
> division knock a few values off balance. But, the result is still
> WRONG and I just don't know enough about roundoff error to know if
> this is an insurmountable problem (but my educated guess is: yes.) Is
> there some clever way to make HISTOGRAM behave properly in such
> situations?
```

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)'
1.99999999999999928946
```

The subtraction and division don't knock a few values "off balance", the values are simply

not representable to the precision you require - such is the nature of float point numberology. As other posters suggested, you need to preprocess the values somehow to make them integers, or take into account the precision when you do the comparisons required. It's worth writing a separate function to do floating point number comparisons. In Fortran I do the following:

```
!      ABS( x - y ) < ( ULP * SPACING( MAX(ABS(x),ABS(y)) ) )
!
!      If the result is .TRUE., the numbers are considered equal.
!
!      The intrinsic function SPACING(x) returns the absolute spacing of numbers
!      near the value of x,
!
!      { EXPONENT(x)-DIGITS(x)
!      { 2.0                for x /= 0
!      SPACING(x) = {
!      {
!      { TINY(x)            for x == 0
!
!      The ULP optional argument scales the comparison.
```

where

```
!      ULP: Unit of data precision. The acronym stands for "unit in
!      the last place," the smallest possible increment or decrement
!      that can be made using a machine's floating point arithmetic.
!      A 0.5 ulp maximum error is the best you could hope for, since
!      this corresponds to always rounding to the nearest representable
!      floating-point number. Value must be positive - if a negative
!      negative value is supplied, the absolute value is used.
!      If not specified, the default value is 1.
```

I don't know off the top of my head how to translate all that to IDL, but it would be a useful thing to do.

paulv

```
> =====
>
> "Nothing shocks me. I'm a scientist." - Indiana Jones
> -----
> Tim Robishaw                UC Berkeley Astronomy
```

Maybe a necessary, but not sufficient, condition to stave off the shocks. :o)

--

Paul van Delst
CIMSS @ NOAA/NCEP/EMC
Ph: (301)763-8000 x7748
Fax:(301)763-8545

Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [meinel](#) on Thu, 12 Jun 2003 15:03:09 GMT
[View Forum Message](#) <> [Reply to Message](#)

Craig Markwardt wrote ...

> timrobishaw@yahoo.com (Tim Robishaw) writes:

> ...

>> Well, there ya go. It's a roundoff error problem that results from
>> trying to balance the values on a razor's edge... the subtraction and
>> division knock a few values off balance.

>

> Partial solution number four: work in powers of 2 instead of multiples
> of 0.05.

Isn't this in the FAQ somewhere?

The technical answer: Computers represent numbers as base 2; humans represent numbers as base 10. Mapping numbers from base 10 to base 2 is exact, mapping numbers from base 10 to base 2 ain't.

$5+5/10 = 5+1/2$ exact correspondence

$5+45/100 = 5+1/4+1/8+1/16+1/128+1/256+1/2048+\dots$ not exact, even in DP

So, depending on the machine precision, 5.45 on the computer is either slightly greater than or less than an exact representation of 5.45. On top of that, neither is 0.05 exact on the computer, so your bin size is also slightly different than what you are expecting.

Bottom line: if you want the results to be exact, think like a machine.

Ed

PS. Let's just say, ...

Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [timrobishaw](#) on Thu, 12 Jun 2003 21:54:13 GMT
[View Forum Message](#) <> [Reply to Message](#)

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)'  
> 1.99999999999999928946
```

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.

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)'  
>> 1.99999999999999928946
```

>
> 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 where I'd be), you could do the integer conversion yourself, perhaps rounding to the nearest bin boundary for values close ($< \epsilon$) to such a boundary, and then perform your histogram on the resultant integer data.

Good luck,

JD

Subject: Re: HISTOGRAM and the Razor's Edge.
Posted by [Paul Van Delst\[1\]](#) on Fri, 13 Jun 2003 13:58:53 GMT
[View Forum Message](#) <> [Reply to Message](#)

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.

Uum... no worries.

> That's not a very friendly way to encourage people to
> post questions to your group.

Oopsy...methinks you misinterpreted my "tone of type". Apologies if my reply offended. I
just meant to point out that the issue was not HISTOGRAM related...but floating point
arithmetic related.

And, please, just because I'm a lousy writer who conveys the wrong message with his usenet
posts, don't let that stop you posting. (Especially about HISTOGRAM.....I still can't
figure out how to use it right :o)

paulv

--

Paul van Delst
CIMSS @ NOAA/NCEP/EMC
Ph: (301)763-8000 x7748
Fax:(301)763-8545
