
Subject: Re: Histogram and bin sizes

Posted by [Kenneth P. Bowman](#) on Thu, 21 Feb 2008 14:54:54 GMT

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

In article

<3a7ec87f-f7e6-40b8-a6de-8328810e0e41@62g2000hsn.googlegroups.com>,
Conor <cmancone@gmail.com> wrote:

> You can always do whatever binning you want, you just have to
> transform your data to the new space and then bin it constantly. Why
> doesn't histogram let you use arbitrary binsizes? Not being an IDL
> developer I don't know for sure, but I would guess it's a speed
> issue. The simpler a program is the faster it is. I use histogram
> all the time because it's one of the speedier programs in IDL. It
> would make me very sad if in order to make histogram more flexible, it
> also became much slower, especially since by transforming my data set
> I can use arbitrary bin sizes for histogram.

If your bins are of uniform width, or if you can transform your data
such that your bins are of uniform width, then it is possible to
compute the bin index for each value with a simple linear
transformation (and conversion to LONG if necessary). This will run at
1 value per clock cycle on some machines.

If the bins are of irregular width, such that no simple transformation
will serve, then is necessary to *search* the list of bins to determine
the bin index. This is necessarily much slower than an arithmetic
calculation, as it involves unpredictable branching.

In this example, binning 10^7 numbers into irregular bins is about 10
times slower than using HISTOGRAM with regular bins due to the time
required to do a binary search.

;Use HISTOGRAM to bin 10^7 numbers with evenly-spaced bins

```
r = 100.0*RANDOMU(seed, 10^7)
t0 = SYSTIME(/SECONDS)
h1 = HISTOGRAM(r, MIN = 0.0, BINSIZE = 1.0)
PRINT, 'Time for evenly-spaced bins : ', SYSTIME(/SECONDS) - t0
```

;Use VALUE_LOCATE and HISTOGRAM to bin 10^7 numbers
;with unevenly-spaced bins

```
bins = [0.0, FINDGEN(99) + 0.1*RANDOMU(seed, 99)]
t0 = SYSTIME(/SECONDS)
i = VALUE_LOCATE(bins, r)
```

```
h2 = HISTOGRAM(i, MIN = 0, BINSIZE = 1)
PRINT, 'Time for unevenly-spaced bins : ', SYSTIME(/SECONDS) - t0
```

```
IDL> @hist_bins
Time for evenly-spaced bins :    0.076565981
Time for unevenly-spaced bins :    0.79658723
```

Ken Bowman
