Subject: Re: Box-Whisker plots in IDL
Posted by jschwab@gmail.com on Tue, 21 Aug 2007 00:06:31 GMT
View Forum Message <> Reply to Message

On Aug 20, 7:47 pm, David Fanning <n...@dfanning.com> wrote:
> jsch...@gmail.com writes:
>> Pardon me if I'm mistaken, but I think these "quartiles with
>> histogram" examples, including the one that's in JD's histogram
>> tutorial are fundamentally incorrect.
>
>> You are assuming "Equal bin widths" ==> "Equal #'s in each bin" !
>
>> When HISTOGRAM splits a data list into N bins, it does so such that
>> the *width* of the bins are equal. In no way does it somehow create a
>> situation in which the *number of points* in each bin is equal (which
>> is what would be required to find quartiles in such a manner).
>
>> The given examples have only "worked" because you're either dealing
>> with uniform distributions (in which case equal bin widths do imply
>> equal numbers in each bin) or because the example data happens to be
>> roughly uniform.
>
>> If you want to convince yourself, try one of those codes with
>> data = randomu(seed, 1000) * 100.
>> and then with
>> data2 = data * data
>> The quartiles in the 2nd case should simply be the squares of the
>> quartiles from the first.
>
> Humm. Maybe you are right. (Isn't it odd that math types
> never hit the SEND button until someone else has made
> a fool of themselves?)
>
> OK, how about this:
>
> data=randomu(sd,100)*100
> minVal = min(data)
> maxVal = max(data)
> medianVal = median(data,/even)
>
> ; Find the quartiles.
> qtr_25th = Median(data[Where(data LE medianVal, countlowerhalf)])
> qtr_75th = Median(data[Where(data GT medianVal, countupperhalf)])
> void = Where(data LT qtr_25th, countlowerquarter)
> void = Where(data GE qtr_75th, countupperquarter)
>
> Print, minVal, maxVal, medianVal, qtr_25th, qtr_75th
> Print, countlowerquarter, countlowerhalf-countlowerquarter, $

> countupperhalf-countupperquarter, countupperquarter
> END
>
> Which gives me:
>
> 1.74060 99.8840 53.5631 31.7422 73.8378
> 25 25 25 25
>
> Cheers,
>
> David
> --
> David Fanning, Ph.D.
> Fanning Software Consulting, Inc.
> Coyote's Guide to IDL Programming:http://www.dfanning.com/
> Sepore ma de ni thui. ("Perhaps thou speakest truth.")

This looks good to me. I wrote a routine to find the energy quartiles of some x-ray data a few weeks back and that was the way I ended up doing it. Not that speed is an issue, but I'd be curious to see how this method compares with a SORT, or some other (yet undiscussed) method. Maybe I'll play around with that tonight if I have some extra time.

Cheers,
Josiah
--
Josiah Schwab
MIT, Course VIII
