
Subject: Re: Box-Whisker plots in IDL

Posted by [teich](#) on Mon, 20 Aug 2007 20:22:56 GMT

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

On Aug 20, 3:40 pm, David Fanning <n...@dfanning.com> wrote:

> te...@atmsci.msrc.sunysb.edu writes:

>> Well, I am looking into the histogram procedure, but I am not getting
>> what I think the 25th and 75th quartiles should be. It seems
>> histogram is not so easy to master. What I am looking into is doing
>> the following:

>

>> data=randomu(sd,100)*100

>> box plot needs min, max, median which are straight forward:

>

>> min(data)

>> max(data)

>> median(data,/even)

>

>> For the quartiles I am trying:

>

>> lower_ind=where(data lt median(data,/even))

>> upper_ind=where(data gt median(data,/even))

>> qtr_25th=median(data[lower_ind(0):lower_ind(n_elements(lower_ind)-1)],/
>> even)

>> qtr_75th=median(data[upper_ind(0):upper_ind(n_elements(upper_ind)-1)],/
>> even)

>

>> However, I think this would work only for a monotonically increasing
>> array. I am not sure how to get 'data' like that. If anyone wants to
>> add to this, feel free.

>

> I calculate it like this:

>

> data=randomu(sd,100)*100

>

> minVal = min(data)

> maxVal = max(data)

> medianVal = median(data,/even)

>

> ; Find the quartiles.

> binsize = (maxVal - minVal) / 4.0

> h = Histogram(data, BINSIZE=binsize, REVERSE_INDICES=ri)

> qtr_25th = Median(data[ri[ri[0]:ri[2]-1]])

> qtr_75th = Median(data[ri[ri[2]:ri[4]-1]])

>

> Print, minVal, maxVal, medianVal, qtr_25th, qtr_75th

> END

>

```
> With 100 values I get this:
>
>      0.401314    98.0063    58.9402    20.0477    73.3419
>
> With 10000 values, I get this, which leads me to think the
> algorithm might be correct:
>
>      0.0249010    99.9960    49.9658    25.0268    74.8059
>
> 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.")
```

Hi, Suppose data is something simple like:

```
data=[2,3,5,7,7,10,11,11,12,15,16,17,17]
```

I get a 75th quartile of 11.0. Shouldn't I get around 15? Maybe I made a mistake. However, what about if I revise what I wrote before to:

```
sorted_data=data(sort(data))
```

```
lower_ind=where(sorted_data lt median(sorted_data,/even))
upper_ind=where(sorted_data gt median(sorted_data,/even))
qtr_25th=median(sorted_data[lower_ind(0):lower_ind(n_elements(lower_ind)-1)],/
even)
qtr_75th=median(sorted_data[upper_ind(0):upper_ind(n_elements(upper_ind)-1)],/
even)
```

Howie
