Subject: Re: compute quartiles of a distribution
Posted by Jeremy Bailin on Wed, 19 Oct 2011 04:32:59 GMT
View Forum Message <> Reply to Message

On 10/18/11 3:48 PM, Jeremy Bailin wrote:
> On 10/18/11 12:12 PM, bing999 wrote:
>> Thanks to both of you for your answers.
>>
>> The procedures in summary.pro and cgBoxPlot.pro compute "real"
>> quartiles. Actually, I should not have used this word in my case i
>> guess.
>>
>> What I want is the interval [M-Q;M+Q] which encompass 75% of the
>> values of the sample around the mean (not the median) value M, where Q
>> is unique (i.e the same at lower and higher values around M). I do not
>> want the 37.5% above M and the 37.5% below. It makes a little
>> difference with what is calculated with your routines.
>> The idea would be to span the sample starting from the mean, and
>> counting the points at lower and higher values around the mean in an
>> iterative manner, until I have counted 75% of sample. This would give
>> the value of Q at which the 75% is reached. I have a crude idea to do
>> that with for loops but it will take forever...
>>
>> If you see what I mean, and if you have a piece of code, this could
>> help a lot!
>>
>> Thanks again.
>>
>>
>>> bing999 writes:
>>>> I have sample of data (which distribution is unknown) of mean M. I
>>>> would like to calculate the quartiles with IDL, i.e what is the value
>>>> of Q for which 25% (or 75%) of the sample is comprised between [M-Q;M
>>>> +Q] ?
>>>> Do you know a routine which does that?
>>>
>>> cgBoxPlot.
>>>
>>> Cheers,
>>>
>>> David
>>>
>>> --
>>> David Fanning, Ph.D.
>>> Fanning Software Consulting, Inc.
>>> Coyote's Guide to IDL Programming:http://www.idlcoyote.com/
>>> Sepore ma de ni thui. ("Perhaps thou speakest truth.")
>>

>
> Easy enough (untested):
>
> data = [.....]
> frac_to_enclose = 0.75
> meanval = mean(data)
> absdiff = abs(data-meanval)
> quartile_index = floor(n_elements(absdiff) * frac_to_enclose)
> q = absdiff[quartile_index]
>
>
> But I share David's concern that this may not really be what you want...
>
> -Jeremy.

Okay, now that I've tested it, there's clearly a SORT missing.
Substitute the last line with:

q = absdiff[(sort(absdiff))[quartile_index]]

-Jeremy.
