

Re: How to Trap Rogue Data ?

Source: <http://www.tech-archive.net/Archive/Excel/microsoft.public.excel.charting/2004-07/0404.html>

From: bill.. (b_at_c.com)

Date: 07/16/04

Date: Fri, 16 Jul 2004 12:15:58 GMT

Thanks for the suggestions

I understand averaging the previous and next points but what is 1-Norm fitting?

Bill

On Fri, 16 Jul 2004 09:51:45 +0100, Martin Brown
<|||newspam|||@nezumi.demon.co.uk> wrote:

>In message <5m2ef0h1c9sgqg81ukmgbdre7u7cjp5cel@4ax.com>, bill..
><b@c.com> writes
>>
>>I have an application that generates hourly system performance
>>logfiles which I graph to look for long term trending.
>>The metric I use gradually varies from 1% to about 15% depending on
>>various external factors – such as time of day and day of week.
>>
>>My problem is that the logfiles sometime hiccup and generate bad data
>>resulting in huge spikes in my curve. I have trapped for the big ones
>>> 20% in my source data but I need something smarter so I can catch
>>large deviations from the curve.
>>
>>Unfortunately I do not have the option to fix the application that
>>generated the bad data.
>>
>>
>>Are there any statistical function that I can use to look for such
>>deviations?
>
>If you are sure they are definitely limited to single spikes on a
>nominal but noisy smooth trend line then the local second derivative
>estimator is a reasonable test. Set a threshold on that to decide on
>rogue points.
>
>ABS(x[i-1]+x[i+1]-2x[i])
>

microsoft.public.excel.charting: Re: How to Trap Rogue Data ?

>You really need to be sure that they **are** rogue though. The test has no
>way of knowing what you really intend. An alternative is to use 1-Norm
>fitting which will safely ignore modest numbers of rogue points.
>>
>>Something that would allow me to toss any data over 5% from the trend
>>line would be perfect.
>>
>>eg
>>
>>value
>>1.2
>>1.9
>>2.4
>>3.1
>>2.6
>>11.3 toss this one using NA() since it is way off the curve
>>3.4
>>4.6
>>6.3
>>7.5
>>9.3
>>11.3 keep this one since it is not too far off the curve
>>8.8
>
>I prefer my plots with all noise displayed and to fix the problem at
>source. You never know when sensor or equipment failure might produce
>real spikes in the signal that a filter will helpfully throw away.
>
>Regards,