

Re: Programmatically finding "significant" data points

Source: <http://coding.derkeiler.com/Archive/Python/comp.lang.python/2006-11/msg01970.html>

- *From:* Jeremy Sanders <jeremy+complangpython@xxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Tue, 14 Nov 2006 14:08:50 +0000
-

erikcw wrote:

I have a collection of ordered numerical data in a list. The numbers when plotted on a line chart make a low-high-low-high-high-low (random) pattern. I need an algorithm to extract the "significant" high and low points from this data.

....

How do I sort through this data and pull out these points of significance?

Get a book on statistics. One idea is as follows. If you expect the points to be centred around a single value, you can calculate the median or mean of the points, calculate their standard deviation (aka spread), and remove points which are more than N-times the standard deviation from the median.

Jeremy

—

Jeremy Sanders
<http://www.jeremysanders.net/>

.