Subject: Re: "Correct" Data Philosophy
Posted by Paul Van Delst[1] on Fri, 18 Dec 2009 18:03:59 GMT
View Forum Message <> Reply to Message

Laura wrote:
> On Dec 17, 11:43 am, David Fanning <n...@dfanning.com> wrote:
>> Folks,
>>
>> Every couple of weeks I get an e-mail from someone whose
>> data is "missing" and they want to replace it with the
>> "correct" value. These e-mails bug me because if the
>> data is "missing" how the hell would I know what the
>> "correct" value is suppose to be?
>>
>> But, generally speaking, they want some method to
>> guess at the "correct" values by looking around the
>> neighborhood, shuffling their feet, etc. I guess we
>> have all been tempted to fudge data, if only for
>> aesthetic reasons, so maybe it is a legitimate request.
>>
>> What would you tell them to do?
>>
>
> Is it similar to "interpolation" or "approximation" or "estimation"?
>
> How about linear/bilinear/trilinear interpolation? Or minimum
> curvature surface or thin-plate-spline? It also depends on how many
> values are available and/or missing. There are other fitting/
> interpolation functions too.

And, just to emphasise the case dependence of an interpolate solution to this problem: do you need the derivatives of your data to be continuous? If so, not just any old interpolation function will do.

Ken Bowman hit it on the head: adapt your analysis and display methods to the data.

Anything else is what I would call "plotology" (is to data display as, e.g., astrology is to astronomy). If I see "real" data (i.e. from some sort of instrument or model/analysis) and it looks beautiful to behold, I'm immediately suspicious.

The "correct" philosophy to have towards data is, IMO, to not have one. The data is what it is. If one expects it to be something else, their analysis will likely trend it that way. That's called bias. Most data is already biased but one hopes it's mainly due to our measurement errors or lack of understanding about the real world rather than our massaging techniques. (bias correction techniques are themselves the subject of many meetings and conferences)

Anyway....

cheers,

paulv