Subject: Re: Correlation on log-log? -or- Easy way of removing 0's?
Posted by noymer on Fri, 26 May 2000 07:00:00 GMT
View Forum Message <> Reply to Message

In article <392E7F7C.9CDEDD62@mathstat.dal.ca>,
  Simon de Vet <simon@mathstat.dal.ca> wrote:

>
> I think that this may be because some of the values are 0, and this
> makes my computer explode (metaphorically speaking, of course). Is
>

I am not sure I have "the answer" to your question, but I have a few
vague suggestions:

1) Often people do log(1+Y) rather than log(Y) to avoid FP errors.

2) If you think in terms of the regression hyperplane rather than the
   correlation coeficient, then the square of your correlation
   coefficient r (called, well, r^2) is the percent of the variance
   explained by the hyperplane (in this case (?) a line). Now,
   think of the slope of the hyperplane let's call it B. For
   raw data B gives change in Y for unit change in X. For log(Y),
   B gives proportional change in Y for unit change in X, and for
   log-log, B is an elasticity. So you are not measuring excatly
   the same thing every time.

HTH,
    Andrew

noymer@my-deja.com

Subject: Re: Correlation on log-log? -or- Easy way of removing 0's?
Posted by Craig Markwardt on Fri, 26 May 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Simon de Vet <simon@mathstat.dal.ca> writes:

> I have been creating scatter plots on a log-log graph, and they look
> pretty good.
>
> However, I'd like some more accuracy than saying "pretty good". I tried
> to use CORRELATE, and got some fairly reasonable values. However, since

> CORRELATE is working with the pure data, scatters that look good on the
> log-log plot can give some pretty horrendous correlation coefficients.
>
> Therefore, I thought I'd try doing a correlation of the log the data, to
> get a better impression of what I'm seeing. Of the five data sets, 2
> gave better correlations, one got worse, and 2 gave me NaN's.
>
> I think that this may be because some of the values are 0, and this
> makes my computer explode (metaphorically speaking, of course). Is there
> an easy way of fixing this problem without loops? Could there be another
> cause of my problem?

This is a classic case of using WHERE to filter your data.

```
wh = where(x GT 0 AND y GT 0, ct)
if ct EQ 0 then message, 'ERROR: there are no valid points'

x1 = x(wh)
y1 = y(wh)

plot, x1, y1, /xlog, /ylog, ...
correlate, x1, y1, ...
etc.
```

Craig

--

---------- ----------
Craig B. Markwardt, Ph.D. EMAIL: craigmnet@cow.physics.wisc.edu
Astrophysics, IDL, Finance, Derivatives | Remove "net" for better response
---------- ----------

Subject: Re: Correlation on log-log? -or- Easy way of removing 0's?
Posted by Alex Schuster on Fri, 26 May 2000 07:00:00 GMT
View Forum Message <> Reply to Message

Simon de Vet wrote:

> I think that this may be because some of the values are 0, and this
> makes my computer explode (metaphorically speaking, of course). Is there
> an easy way of fixing this problem without loops? Could there be another
> cause of my problem?

To use correlate() with all positive values only, you can do something
like this:

```
index = where( array1 gt 0.0 and array2 gt 0.0 )
result = correlate( alog( array1[index] ), alog( array2[index] ) )

        Alex
--
  Alex Schuster     Wonko@weird.cologne.de          PGP Key available
                    alex@pet.mpin-koeln.mpg.de
```