

Re: linear regression results approximates to the mean of Y

Re: linear regression results approximates to the mean of Y

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2008-06/msg00114.html>

- *From:* Greg Heath <heath@xxxxxxxxxxxxxxxxxxx>
 - *Date:* Thu, 5 Jun 2008 02:46:45 -0700 (PDT)
-

On May 29, 3:06 am, feilian <bs...@xxxxxxx> wrote:

I do a linear regression on a data set $\{(X_i, Y_i) | i=1, \dots, n\}$ and get a linear function $Y=AX$;

You could use

$$(Y - \text{mean}(Y)) = A * (X - \text{mean}(X))$$

but it is better to use

$$Y = A * X + B$$

then, when using a test set, I find that the mean of estimated results always near $\text{mean}(Y)$.

If the test (out of sample) set is drawn from the same probability distribution as the training (in sample) set, this is what you would expect if both sets are sufficiently large. After all, both sample means are estimates of the population mean.

Does this mean the X and Y have no linear relations?

No, and I don't understand why you would think so.

I don't know whether I expressed clearly?

Your confusion can be minimized, if

1. you always include an intercept term in your model

Re: linear regression results approximates to the mean of Y

Re: linear regression results approximates to the mean of Y

2. You apply your model to data that is assumed to come from the same distribution as the training data used to estimate A and B.
3. Both sets are sufficiently large

Hope this helps.

Greg

.