

Re: Linear regression

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2006-05/msg00843.html>

- *From:* "Reef Fish" <Large_Nassau_Grouper@xxxxxxxxx>
 - *Date:* 31 May 2006 11:05:01 -0700
-

Jens wrote:

Thanks Bob for your prompt answer.

<skip>

That's the usual assumption about the ERRORS of the conditional distribution of Y given X. More specifically iid $N(0, \sigma^2)$ is the usual assumption.

<skip>

There's nothing to investigate until AFTER you've fitted some tentative model. THEN, you should examine the residuals (observed errors) to validate the iid assumptions.

So my model should be $Xb=y+e$ and $e=Xb-y$ should be "close to" iid $N(0, \sigma^2)$.

It's $Y = Xb + e$, where e is nearly iid $N(0, \sigma^2)$.

$Y = Xb$ is what you try to model or fit. What does NOT fit is lumped into the random error term e . e may be the results of hundreds and thousands of variables that actually have SOME effect on what you measure, but you simply (for simplicity sake) assume it comprises ONE single random error.

<skip>

By "couplings" I think you mean either statistical or deterministic relations some pairs of X's. Short answer, as long as the X's are linearly independent, there's nothing to worry about.

Re: Linear regression

By couplings I mean some unknown relations between the columns. Unknown in the sense that no-one has investigated such relations properly, but based on "intelligent" guess some relations surely exist.

But you're NOT interested in those unknown, uninvestigated between the "columns" (on the RIGHT hand side — the independent variables). If you want to explore its relation with the others, put it on the LEFT as the dependent variable.

<skip>

You seem to have the common misconceptions and misunderstanding between "linear independence" (linear algebra) and stochastic (statistical) independence of the error terms. There are several threads in sci.stat.math specifically about the difference of these two concepts.

Yes indeed I need to try to understand this topic a better manner. Actually my current observations are only a subset of a larger set. I'll return later with more specific questions.

What you are asking are the most BASIC questions regarding a Linear (Regression) model. The fact that you wrote your model as

$$Xb=y+e \text{ and } e=Xb-y$$

suggests that you haven't read any book on the subject. Try Neter, Wasserman, Kutner & Nachsheim's book on "Applied Linear Models" which is a standard textbook in many universities and recommended by many (including myself) as a good book on the subject.

— Bob.

.