

Re: large negative parameter correlations in regression

Source: <http://sci.tech-archive.net/Archive/sci.stat.math/2006-09/msg01064.html>

- *From:* "Reef Fish" <Large_Nassau_GrOuper@xxxxxxxxxx>
 - *Date:* 26 Sep 2006 10:07:49 -0700
-

Stephen Clark wrote:

I have a simple 3 variable plus intercept regression model with good goodness of fit statistics – high R^2 and t -ratios.

When you say you have "simple 3 variable plus intercept" regression model, do you mean just a multiple regression of Y on three X 's with an intercept, or do you mean something else?

When I calculate the correlation matrix of the parameter estimates I get a large negative correlation between the intercept and one of the other parameters (-0.98).

Did the correlations come from the Covariance matrix of the estimated betas? That is, is your correlation matrix derived from

$\text{inverse}(X'X) * s^2$

where X includes the column of 1's and your three X variables, and s^2 is the variance of the residuals?

That would be the Covariance matrix of your estimated (b_0 , b_1 , b_2 , and b_3).

If your negative correlation -0.98 came from that Cov matrix, then it would mean one of your estimate b_i is negatively correlated with the estimate of b_0 . It is related to the multicollinearity concept but not a measure of multicollinearity.

To investigate whether I have multicollinearity I have used the R-package and the VIF is only given for the non-intercept parameters in the model (the

Re: large negative parameter correlations in regression

VIFs are 1, 7.5 and 7.5).

Leave the packages out of any such discussion because unless you can describe exactly what the package computes and delivers, one can never be sure what you are talking about.

Typically VIF refers to the reciprocal of $(1-R^2)$ where R is the multiple R of one X regressed on the remaining X's in the regression.

If so, then your VIF simply means that the max R between any one of the X's with the other 2 is at most .93. hardly an alarming factor.

There are no worrisome correlations in the ACTUAL variables, including a constant intercept.

What do you mean "worrisome correlations in the actual variables"? :-)

How do you know? Where are your partial correlation information? (That is in part imbedded in the significance of the T's)

Should I be concerned? Thanks.

My concern is your throwing these terms around without apparent theoretical understanding of what they are telling you. "RULES OF THUMB" in any regression analysis are dangerous toys to play with. It is easy to become habit forming and become all thumb.

I would suggest you pay little attention to what Bob O'Hara says in Regression matters. He is full of thumbs and a frequent Quack on that subject.

-- Reef Fish Bob.

.