Review of: "Decoding the Correlation Coefficient: A Window into Association, Fit, and Prediction in Linear Bivariate Relationships"

Xinmin Chu[1]

1 University of Massachusetts Boston

Potential competing interests: No potential competing interests to declare.

After carefully reading this article and all the posted comments till now, I think, this is an initial attempting manuscript with critical-thinking-style exploration to a fundamental and popular (even trivial, as commented once) statistical measure between variables, other than a thoughtful, publishable paper. It is true that the mathematical statements, proofs, and organized paragraphs are quite important for academic research, that should be taken into consideration by the author.

But, from the sentence in the article saying, "Even without conducting regression analysis, the correlation coefficient (r) can still provide an estimation of prediction, although not always accurately and reliably", I prefer to considering the correlation coefficient (r) as a black-box that the author might be trying to break it out to pioneer a new path to "prediction without regression" with this starting work.

Based on this, my suggestion is listed as following.

Focus on some related equations in Rodgers and Nicewander (1988) - as listed in the reference - and find good proofs and direction to innovative exploration.

For instance, from equation (3.1), we have

$$r = b_{Y \cdot X}(s_X/s_Y) \Rightarrow b = r(s_Y/s_X))$$

plug it into the linear bivariate relationship,

$$Y = a + bX + \varepsilon = a + r\frac{s_Y}{s_X}X + \varepsilon$$

$$\Rightarrow \varepsilon = Y - a - r\frac{s_Y}{s_X}$$

through minimize the error (ε), we can get an estimated correlation coefficient ($\hat{r}$).

Meanwhile, it is necessary to include assumption on the ratio of standard deviation (or standard error). Then, conduct elaboration and discussion on the possibility / capability of using correlation coefficient for prediction, and the advantage, if existing, comparing with the performance using slope (b) in traditional statistics by simulation.

For the variation (r to r^2), take a start from the equation (5.1) in the part 5,

$$r = \frac{s_{\hat{Y}}}{s_Y}$$

figure out the relationship and discuss the application of correlation coefficient, and so on so forth.

Not sure but worth discussing, using r in prediction could have advantage when it is more difficult to estimate the slope (b) than to estimate the correlation coefficient (r), such as, for higher-order data (tensor).

In one word to this article, it is a good try, and it is worth ploughing into seemly solid ground.

Hope this is helpful.