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

Fabio D'Ottaviano

Abstract:

The abstract fails to clearly define "r" as the Pearson Correlation Coefficient, among other correlation coefficients that exist. There is confusion between the coefficient of determination (R^2) and correlation (r). At one point, the discussion shifts from "r" to the coefficient of determination without mentioning R^2, creating ambiguity in the topic under discussion.

Introduction:

The author seeks to address a problem that is essentially a result of poor methodology: using the correlation coefficient for prediction. This is not a standard practice among those more skilled in the art, hence the objective of the study is somewhat unclear.

Argument:

Only in the sentence "Upon reflection, it seems contradictory to claim that r squared demonstrates how much...", is R^2 referred to as the coefficient of determination.

In the statement "Instead of stating that a higher coefficient of determination should never be interpreted as indicating stronger conservatism in older people compared to younger people (in comparison to a lower coefficient), it would be more accurate to say that "it may not necessarily always imply that..."", the example lacks sufficient details of the scenario being discussed, making it hard to follow the argument here and when it's referenced later.

In the assertion "It is important to remember the relationship between the coefficient (b) and the correlation coefficient (r), where b can be replaced by r times the ratio of the standard deviations (SDy/SDx), whether standardized or not. Therefore, a higher coefficient (b) corresponds to a higher correlation coefficient (r).", the conditions needed for the statement to hold true aren't fully described. It should be clarified that (SDy/SDx) must remain constant for b to represent r and vice versa - if this is what was meant.

In the claim "When the SD of X is much larger than the SD of Y, it indicates that the cases are closer to a nearly horizontal line, resulting in a smaller slope and a higher correlation coefficient.", again, the conditions under which this statement would be valid are not fully stated.