Reviewer's report

Title: Spurious interaction as a result of categorization

Version: 0 Date: 28 Sep 2018

Reviewer: Brian Tom

Reviewer's report:

This is a well written paper demonstrating once again that care should be taken when trying to decide whether it is appropriate to dichotomize/categorize continuous variables when investigating regression effects (in this case, interaction effects) in regression models. The novelty of this paper was demonstrating that categorization may lead to a spurious interaction (qualitatively) in multiple regression models. Precise analytical expressions in the multiple linear regression case are provided for the situation where the two independent variables are bivariate normal.

I have a number of comments which I hope will improve the paper.

Illustrated examples: The author in both examples demonstrate that different findings can result from treating the independent variable(s) as either continuous or binary and this is fine. However, from these results the author conclude that "Both examples show how categorization may lead to apparent interactions in practical data analysis". This statement is somewhat implicitly assuming that the continuous independent variables versions of the regression models are "correct" and that the dichotomized independent variables versions are incorrect. However from what has been presented we do not know which of these models are mis-specified and in fact all of them could be mis-specified. I think the author therefore needs to be careful with the conclusions drawn from these illustrative examples, unless some more detailed modelling is presented (e.g. regression models with say splines to model the functional relationships between the independent variables and their interaction with the outcome).

Results: In the paper, the emphasis is on the estimated interaction effects and not on the standard errors of these effects. It may be the case that the estimated strength of the interaction effect (from the truth of no interaction effect) will change (say increase) with the dichotomization of the independent variables, but the test for the interaction effect may still result in non rejection of the null hypothesis of no interaction effect because the standard error also increases. I therefore feel that the author should also incorporate into the article more discussion of the impact the estimated standard error may have on inference, as both the magnitude of the estimated regression coefficient and the standard error of the estimate will determine whether a decision of statistical significance is arrived at or not. I also will like to highlight the article written by Farewell, Tom and Royston (2004) in the Journal of American Statistical Association which discusses "The Impact of Dichotomization on the Efficiency of Testing for an Interaction Effect in Exponential Family Models".
Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Unable to assess

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics

Quality of written English
Please indicate the quality of language in the manuscript:

Acceptable

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organization that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?
If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

'I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal