Responses to reviewers

Ms. No. ESPR-D-15-00741
submitted to
Environmental Science and Pollution Research

Eduard Szöcs and Ralf B. Schäfer

March 11, 2015

Dear Dr. Schulz and reviewers,

We are thankful for reviewing our manuscript and the comments that helped to improve the paper. We revised the manuscript according to comments of the three reviewers and are re-submitting the manuscript for consideration for publication in Environmental Science and Pollution Research.

In the remainder of this document, we describe the changes that we have made to the paper for resubmission. To assist the assessment of our changes we have submitted two versions of the revised manuscript: one with highlighted changes and another without any highlighting.

Kind regards,
Eduard Szöcs and Ralf B. Schäfer

Response to Reviewer 1

Response to Reviewer 2

Comment: One point to consider in a resubmission is that GLMs aren't for use for all non-normal data - some qualifications are needed of when to use GLMs. Distributionally, GLMs replace the assumption of normality with the assumption that data come from the exponential family, which does not cover all situations. More critically, GLMs replace the assumption of equal variance with an assumption that the mean-variance relationship follows a pre-specified pattern, and violation of this assumption has similar impacts to violations of the equal-variance assumption in a linear model would have. Particular situations where GLMs are useful are binary data (known to be a quadratic mean-variance), count data (especially count data with lots of low counts, which typically cannot be transformed to normality effectively), and proportions constructed from counts. These are the examples highlighted in the text, and are of clear relevance in ecotoxicology, but the argument does not extend much more broadly than these examples - it is not a case of "your data aren't normal so you should use GLMs". It is more accurate to say "if you have binary, counts or proportions from counts you should use GLMs". I think this requires a modest adjustment to the phrasing of the abstract and introduction to better qualify the situations where GLMs are appropriate. e.g. where O'Hara and Kotze and Warton & Hui are mentioned, this should be qualified to refer to counts and proportions (of counts) respectively. I don't think the title needs changing, sufficient qualification elsewhere of the type of non-normality if interest should do the trick.

    Response:

Comment: In the first paragraph, "positive" is used where "non-negative" is needed (i.e. zero is a possible value), also change bonded to bounded

    Response:

Comment: Page 2, second column - the parameterisation isn't equal to the Poisson model - the mean model is the same, I think that is what was meant.

    Response:

Comment: Page 2, second column - also, strictly speaking, not all the methods mentioned here are GLMs. Negative binomial regression ("NB2") is a GLM if the overdispersion parameter is fixed, otherwise it is a generalisation of GLM. Similarly, quasi-Poisson isn't really a GLM, but a related method.

    Response:

Comment: Section 2.2 - a description of what is meant by binomial data would be helpful at the start of this section. Perhaps for completeness a description of what is meant by counts would help at the start of section 2.1

    Response:

Comment: Section 2.2.2 - it would be useful to also mention GLMMs with a random intercept term. Often with this sort of data there are extra sources of variation from one sample to the next that are not explained by the binomial assumption, and this is the most natural way to account for it.
  Response:

Comment: Sections 2.1, 2.2: references to relevant articles or texts would be useful for GLMs, where the reader can find more details at a level appropriate for readership (e.g. a Zuur text? Quinn & Keough?). Also point readers to supp 2 for code and a worked example.
  Response:

Comment: page 3, line 58, the parametric bootstrap
  Response:

Comment: Page 3, second column, line 30: Wald tests should be used with caution, when a mean estimate is zero (or close to it) they can have strange behaviour (related to the issue of separable data - parameter estimates and ses diverging, test statistics going to zero).
  Response:

Comment: Page 4, second column, lower means: it is noted that the means are lower when data are transformed. This is known as transformation bias, and occurs because the transform linear model is actually estimating something different - it is no longer trying to estimate mean response. It would be helpful to say this and maybe include a useful reference. This also gets at the issue of interpretability - a disadvantage of the transform approach is a loss of interpretability, because we are no longer modelling mean response.
  Response:

Comment: page 7 first column, line 7: bounds not bonds
  Response:

Comment: Section 4.1 the case study discussion was a little confusing. So is the point that you can get different results by linear models according to transformation, but that GLMs can resolve the problem of choice of transformation? (by replacing it with a decision about mean-variance relationship?) I think a stronger point is that the GLM is usually a better fit to count data, no transformation can make data Gaussian when it has lots of zeros and small counts, but GLM is designed for such data. If diagnostic tools support this argument for your case study they could be included.
  Response:

Comment: page 7 column 1 line 27: it is not so much unreliable and biased - they are estimating a different parameter (in an unbiased fashion) - they estimate the mean of log(Ax+1) data, and a problem is that this has no natural interpretation in terms

3

of the original data. It is fair to make the point that it is biased as an estimate of the mean response, but it would be helpful to explain why and how the estimators have problematic interpretations.

Response:

Comment: page 7 column 1 line 50: a priori not a priory
page 7 column 2 line 37: data are (data being the plural of datum)
page 7 column 2 line 42: GLMs have (or GLM has)

Response:

Comment: page 7 column 2 line 44: the higher power in Fig 5 seemed to be through the GLMs being more conservative, rather than due to greater efficiency. This issue could likely be fixed using a parametric bootstrap, as was done previously in figure 2. A related point is that the linear model had the advantage that it seemed to be better at maintaining nominal significance levels at small sample sizes, which should be mentioned. As the authors say however it did not have as good power (and also loses something in interpretability and biased estimation of means). Hence while the GLM has a number of advantages, combining it with param bootstrap at small sample sizes might be worth considering to address the Type I error issue.

Response:

Comment: Supplement 2 is a useful addition, and it might be worth giving it a little more emphasis in the text to ensure ecotoxicologists wanting to try out new analyses give it a look. A couple of suggestions though: - there is no parametric bootstrap code there at the moment. This would be easy enough to do by using the mvabund package, calling manyglm (even on a univariate response) then anova(..., resamp="monte.carlo") will do a parametric bootstrap. - you can construct residual plots for GLMs on this package, using the plot function on a fitted manyglm object.

Response:

Response to Reviewer 3

Comment: With its simulations in an ecotox. context, the paper is a fine contribution to make researchers aware of the possibilities and advantages of a generalized linear models (GLM). What is lacking is pointing to the disaster of using GLM in an improper way, notably without adjustment for overdispersion. The paper and abstract can still be made more useful by given not just the recommendation for GLM but also to give advice and warnings of proper usage. For example, (almost) never use loglinear/Poisson regression without adjustment for overdispersion, be aware that a nicer looking model like the negative binomial may give inflated type I error, a case that is overcome in this paper by bootstrapping. The same applies to binomial/logit regression unless the experiment really consist of independent observations for each 0/1 result (and is not just a count with a predetermined maximum). In this light it is strange that the logistic/binomial case is treated differently from the loglin-

ear/Poisson case! Should this not be repaired or at least given more warning. (1) Also, if the authors stick to the binomial/logistic without overdispersion, the normal model could be made powerful in the simulations by treating the error variance as known, because it is a known constant for the arcsine transformation. (2) Also, the simulation now use slight overdispersion whereas the example /typical data has large overdispersion. Add simulations with large overdispersion. The current ones can go to supplementary. (3) Without the use of a GLM equivalent of the Williams test all the advantage of the use of GLM in terms of power are gone. See the example. Discuss this ambiguity. You can perhaps use a bootstrap test based on (GLM?) monotonic regression or similar. I know some cues/leads in this direction.
Response:

Comment: Page 1 Line 22 right column: a continuous proportion is fine for area cover and so on, but not for proportion of surviving animals where is just "k out of n" turn into a proportion. Discrete therefore.
Response:

Comment: Line 58 right add ref to (Warton 2005).
Response:

Comment: Page 2 Line 40 : , for y¿0 should for y¿= 0 and be on the next line without an indent after formula (if it is not new paragraph).
Response:

Comment: Line 51 x_i is undefined. The model appear to specify a simple linear regression or a control-one_treatment model. Neither is used in the paper!!! Modify the notation therefore..
Response:

Comment: Line 59 Mention backtransformation here, which works fine for the confidence intervals, and not the backtransformed is the median on the original scale. If you want to backtransform to the original mean, use exp(mean+ 0.5*sigma2). (Aitchison & Brown 1969) page 8 (Jongman et al. 1995)page 19.
Response:

Comment: Page 3 line 17 n= 4.10 = 40 what does the mean here/ what is the purpose? Line 41 : give more attention to the condition for the binomial: independent observations of each of the successes (k) out of the number of experiments (n). And: thus give attention to overdispersion in this case as well, particularly as you found nasty things for the LR of NB. Line 28 right. Ref for Dunnett contrasts.
Response:

Comment: Line 59 right. Only slight overdispersion, whereas the example data have large overdispersion. Add large overdispersion.
Response:

Comment: Page 4 L 19 (missing. Line 56 left to 2 right. Four p-values: what to believe? Your simulations (although with small overdispersion) show it. The LR of NB cannot be trusted. So the 0.016 is out. The remaining tests all show about the same p-value if you interpret statistics properly. Please note that the distinction between significant and non-signicant (here p = 0.061 and 0.042) in itself is statistically insignificant. (Gelman & Stern 2006).
Response:

Comment: Line 4 right : the backtransformed mean gives a median on the original scale, a value that is more interesting than the mean (generally for skew data), and for skew right always smaller than the mean.
Response:

Comment: Section 3.2 Always start with the Type I error results because the power results are without meaning if the Type I error is inflated. This also applies the order of the figs: interchange the two rows of subfigs.
Response:

Comment: 21 right: remove GLM_nb in this sentence as GLM_nb must be discarded due to inflated Type I error.
Response:

Comment: L32 right. ", but this." Although you should reorder the results, I mention that the "but this" should at least be "This could fortunately be ..."[and we could not have known for sure (apart from the general warnings about LR) without the simulations] This inflated error for GLM_nb must receive more attention. It is an interesting result of this study: you must to bootstrap when using NB.
Response:

Comment: Page 5 Fig. 2 legend. Add that n =100 simulations. Add the interpretation that GLM_nb has inflated type I error, and that its line for power is just added for completeness and should not be interpreted as an estimate of the true power.
Response:

Comment: Page 6 Fig.5 LM for 0.8 has a high Type I error. Does it deviate significantly from 0.05? [same question for the high Type I for GLM_nb in fig 2]. Use confidence interval for binomial p = 0.05 n = 100.
Response:

Comment: Case study : You can check whether the difference between log(2y+1) and log(Ay+1) with A = 0.182 already causes the difference. (it probably does not). The important difference is the use of the Willems test! Without the use of a GLM equivalent of the Williams test all the advantage of the use of GLM in terms of power are gone! See the report of the Williams test below.
Response:

Comment: Page 7 L35 Move ref to Warton to after data. L45 Add a reference to (ter Braak & Šmilauer 2014) who advocate the use of the transformation approach in a multivariate (dimension reduction) context.
   Response:

Comment: Line 1 right (like GLMs) -¿ (like GLMs and the Williams test)
   Response:

Comment: Line 10-40 Again: discuss Type I error first. Then no mention of GLM_nb in power.
   Response:

Comment: Line 18 right. Add after "non-normal data' the point the importance of proper accounting for under and overdispersion when using GLM. What is lacking is pointing to the disaster of using GLM in an improper way, notably without adjustment for overdispersion.
   Response:

Comment: Line 41 right size of 9 ->size of 6-9 (?)
   Response:

Comment: Line 44 right the transformation ->the LM
   Response: