

> -- Editor --

> Thank you for your submission titled "Testing the equality of normally distributed groups' means under unequal variances by doex package" to the R Journal. Our apologies for the delay in processing the article. We would be interested in publishing the article if it is satisfactorily revised.

> We would appreciate a revised version and point-by-point response to the reviewers comments within 2 months. Remember, that when responding to the reviewer's comments, your job is to persuade me, the editor, that either you've dealt with the issue, or that it's not relevant. To this end, please produce a single document that includes all the reviewers comments mingled with your responses. I particularly recommend the strategy described at <http://matt.might.net/articles/peer-review-rebuttals/>.

We revised our manuscript according to the thoughts of referees. Our comments about revisions are given related to the strategy you recommended as follows. Also, the revised parts of the manuscript are given as red color, and the language revisions as green color.

> -- Reviewer --

> Here my comments (in the form of bullet points) to the manuscript 'testing the equality of normally distributed groups' means under unequal variances by doex package'. Overall, the goal of grouping several tests of heteroscedasticity in one package is commendable. The English level of the manuscript needs thorough revision.

The manuscript is checked by Lecturer Edward MCQUAID (<https://akademik.anadolu.edu.tr/emcquaid>), who is an English Native speaker, from Anadolu University, School of Foreign Languages.

> I checked the R code and I feel what's missing (and this needs to be implemented in the package help files) is a graph for the four groups being compared. For this purpose a simple boxplot with notches will do. In the example the authors provide, the notches went outside the box as the sample sizes are too small (~6). I'd suggest the authors provide an example in which sample sizes are larger.

We think that visualization is very valuable in reporting. Since the doex package can be supported by boxplot() commands included in many basic packages in R such as {graphics} in this regard, we do not need to include it in the package. However, we will consider this valuable suggestion in our future studies.

We added an extra real dataset example is "Example 2" to handle this drawback. Also, the boxplots are added to complement the EDA in each of the examples.

> There is no reference in the manuscript as to whether samples need to be independent or not. This issue needs to be made explicit.

We agree on this issue and have made some additions to clarify in the beginning of “Introduction” section. And also, it is indicated that the groups must be independent in the manuscript.

> I reckon the authors should streamline the way they called the tests. The tests are ‘tests for the equality of group variances’; I suggest sticking to this terminology throughout.

We thank the reviewer for raising the point. We stick to this terminology and revised “heteroscedastic ANOVA tests” to “tests for equality of normal distributed and independent group means under unequal variances”.

> I wonder why the well-known Levene’s test wasn’t included. This test is default in SPSS so I reckon it needs to be added to the package (the authors do include though the Brown-Forsythe test and this test is also used in SPSS). I reckon the authors should stress that ‘doex’ allows the user to have access to tests not available in other packages (e.g. SPSS). Hopefully this will appeal to SPSS to turn to R. There are some readings that need to be considered. For example: “Combination of Levene-type tests and a finite-intersection method for testing equality of variances against ordered alternatives” (by Noguchi and Gel, 2010), “A Robust Test for Checking the Homogeneity of Variability Measures and Its Application to the Analysis of Implicit Attitudes” (Erps and Noguchi, 2020)

The reason we did not include the Levene test in the package is because it is included in many R package such as car, rstatix, lawstat, inferr. With this study, we want to stick with the idea of creating a package that includes tests not included in the R ecosyste. We added an explanation about this in the “using the doex package” part. It is the sentence is “After the explanatory data analysis, ...” in page 7. We considered and added the references which are recommended.

We agree that it is emphasized that doex allows the user to have access to tests not available in other packages. To do this we add some expressions in the “Introduction” and “using doex package” parts.

> From tables 1~3, and in the text, it isn’t clear if ALL samples were of normal distributions. If so, this needs to be made explicit everywhere. Also, even if normal distributions and other distributions were used, their parameters need to be reported.

We agree this contribution and the necessary informations about the distributions and parameters of the samples in Tables 1-3 were added and stated clearly.

> I also recommend the authors make their simulation R code available via a repository (e.g. Figshare or GitHub) and provide the link within the manuscript.

We thank the reviewer for raising the point. The codes are added to GitHub and the link is given in the manuscript.

> If the researchers used only normal distributions, such limitation needs to be stressed; ideally addressed. In practice, it’s rather rare to encounter normally distributed data and these papers

evidence so: “Skewness and Kurtosis in Real Data Samples” (Blanca et al., 2013), “Non-normal Distributions Commonly Used in Health, Education, and Social Sciences: A Systematic Review” (Bono et al., 2017)

We only used the normal distribution because we realized that there has not been such an extensive study in terms of the number of the considered tests in the literature. Of course, simulations can be extended for different values of skewness and kurtosis, but this is outside of our focus. We add a phrase to discuss this issue in “discussion” part. It is in the sentence “It is rather rare to encounter normally ...” in page 14. Thanks for this suggestion.

> In relation to normal-samples-only for the simulations: I strongly recommend the authors report the case in which 3, 5 and 7 groups come from, e.g., two-parameters Gamma and/or three-parameters Ex-Gaussian distributions. I suggest any of these two or any other pair of distributions in which their number of parameters don't lead to very complex combinations. I understand using the normal is practical as it's got only location and scale parameters but the Gamma and the Wald distributions are 2-parameters distributions that can be set to have large skewness (e.g. IG $\sim 1, 0.2$).

We thank for this comment. However, the tests we consider were proposed for only normally distributed populations. For this reason, we think that it is a better solution to develop new parametric tests for the distributions you suggested. This idea is in our further studies. We explicit this in the sentence is “It is always optimistic idea to encounter ...” in page 15.

> The authors do type the following “The highly skewed Gamma distribution and Gamma distribution with shape close to being normal are considered.” but it isn't clear if this happened in their article or elsewhere. At any rate, the authors need to be very specific as to the distributions used and their parameters. The reason I stress on providing the case of non-normal examples is simply because applied statisticians will immediately know normal samples aren't the norm and will want to see the power of the tests under realistic cases; i.e. the samples are skewed.

We noticed that we are using a confusing phrase here, so we extracted the phrase “... which are near to the normal distribution to better linguistic expressions”.

> The authors acknowledge that “In further studies, it is planned to expand the package by adding methods used to test the equality of the log-normal and inverse-Gaussian distributed groups' means which are near to the normal distribution.” but I'd say that to appeal users that know normal data aren't typically found, simulations for both of them are carried out and the results reported.

We updated the plans on our future studies and explained it more clearly. What we want to talk about here were tests with complex algorithms developed for versions with nuisance parameters. Some references are given to point out them in the sentence is “Thus, it is planned to expand ...” in page 15.

> Overall, it's a good work but it needs extra fine-tuning to make it more interesting to the user.

Thanks for the contribution of the referee.