

Response to reviewer comments for “htestClust: Hypothesis Tests for Clustered Data under Informative Cluster Size in R”

We thank the reviewer for their thoughtful and insightful comments. We’ve made changes in response to these critiques that we feel have improved the paper, and appreciate the opportunity to resubmit our manuscript to R Journal. Our responses to reviewer comments are below. Reviewer comments are **boxed** and our responses are *italicized*.

Reviewer 1

Overall I think the article is in very good shape. My primary comment is about the title, abstract, and introduction. The type of data collection you are talking about seems relatively common (I've certainly encountered it before) but I don't think the term "clustered data" is particularly well-known. Even when performing a quick Google search about half of the results are about a different topic. When I first skimmed the first page of the paper I thought this was about cluster analysis. In other words, trying to do some kind of analysis after performing an algorithm such as K-means.

I think it would greatly help avoid disappointed readers as well increase the discoverability of this paper/package if a gloss of what is meant by clustered data could be included in the abstract and the first sentence of the article. It certainly becomes clear after reading the whole first page, but I worry that some of your audience won't get that far and will think this is a machine learning paper rather than fundamentally about inference.

We have revised the title of the article, added a sentence to the abstract, and rewritten the opening paragraph of the introduction in an effort to provide more clarity about the type of clustered data this paper refers to. We recognize the possibility for user confusion, but feel it important to maintain terminology by which to refer to these types of data. In the informative cluster size field, the usage of “cluster” is fairly ubiquitous.

As another note, the paper states that "No other R package includes the broad range ... that is provided by htestClust" (page 4). I don't have a counter example to this claim and it quite likely true. However, it is impossible to know exactly what every other R package provides, particularly if we truly consider all packages, such as those in a GitHub repo but not pushed to CRAN. I don't think anything is gained by the wording of this comment and I would recommend softening it a bit ("the authors know of no other package that ...").

We have revised the language as suggested.

The package was easy to install and each of the functions as well documented and included a simple, easy-to-follow example. It would be great to include a vignette with the package; perhaps a slimmed down version of this article could become that vignette.

We have added a vignette to the latest version of htestClust published on CRAN.

In terms of content, the one thing I would like to see in the package are some unit tests. It seems fairly easy to write at least a few sanity checks that make sure the results are the same compared to the classical functions when there is only one item in each cluster (or, the values in each cluster are all identical).

We have added a paragraph to the manuscript noting the general correspondence between htestClust functions and their classical analogues should all clusters have a size 1 (the correspondence is not exact since the tests implemented in htestClust are asymptotic in scope). We have included an example illustrating this result to the manuscript and also noted the correspondence in the help file for the proptestClust function.

On a stylistic level, I would have recommended a different naming convention for the **htestClust** functions. The authors took the classical test functions, removed periods and/or capital letters, and then added the suffix 'Clust' at the end. When looking at the table in the paper the correspondence is clear, but thinking about this from a user's perspective it would be easy for discovery to each name them with a prefix (which can be used to find all the functions with autocomplete) or to retain the exact original name with a new suffix (then autocomplete will show the second option whenever typing the first). I would not require the authors to make one of these changes, but offer it as a potential improvement if they are open to the idea.

We thank the reviewer for this insightful suggestion and we do see the merit for such function nomenclature. We currently prefer to keep our functioning naming convention as they are. Several of our functions behave differently than their analog function, particularly in terms of the arguments they accept. For example, proptestClust accepts raw data (presumably from a data frame) or aggregate data, while prop.test accepts only aggregate data. As such, the analyst looking to use proptestClust in a way similar to prop.test will be stymied. Moreover, the theory behind the reweighted tests in htestClust is detached from those behind the classical hypothesis tests. For these reasons, we prefer to parallel nomenclature while keeping a level of distance from the classical hypothesis test functions. We believe our current naming convention maintains such a balance. That said, we do recognize the end-user convenience of having names as the reviewer has indicated, and will consider this suggestion further for a future major update version to htestClust should user feedback suggest the current function names are burdensome.