The Concordance Test, an Alternative to Kruskal-Wallis Based on the Kendall-τ Distance: An R Package

Detailed Response to Reviewers

The authors appreciate the positive criticism that the reviewer has made on the manuscript. We will deal with the reviewer suggestions next (presenting our response in blue). In the manuscript, the changes are also shown in blue.

Response to Reviewer #1

Overview

This paper presents an R package that implements the authors' new method for computing a non-parametric test for detecting the relationship between multiple groups of numeric or ordinal data. The method is a generalization of the Kendall rank correlation coefficient when we have more than two samples. The implementation of this method is non-trivial because it involves some computationally simulations. The general range of the applications for this approach will make this an interesting paper the R community.

The authors thank the reviewer for the positive assessment of the work.

Below are a few relatively minor points that I think can improve of the article and package. There is one very important question about the package's results compared to the implementation in the base R package; hopefully there is an easy explanation.

Article

I generally found the article to be well written and easy to follow. I particularly liked the historical perspective on non-parametric techniques as well as the gentle (but not overly verbose) general introduction to this class of non-parametric estimators.

We thank the reviewer for the comment.

There are two additional things I think that would be useful to include in the paper; these could be combined into a new short section or sub-section. First, the authors seem to only be presenting simulated data. Or at least, the data that is given has no specific source. I think it would be much better to include at least one real example of the package used on a dataset. Secondly, the stats package in R does include a 'kruskal.test'. It would be very useful to show an example of how this package's functions can be used to replace code that uses 'kruskal.test'; in other words, code that uses the former function followed by code using 'CT_Hypothesis_Test'. You could probably do both of these together by using the Hollander & Wolfe (1973) dataset referenced in the 'kruskal.test' example.

The authors thank the reviewer for the suggestions that have helped us to improve the paper. In the ConcordanceTest package, an example with the Hollander & Wolfe (1973) dataset had already been included in the examples of the function 'CT_Hypothesis_Test', following precisely what is done in the 'kruskal.test' function of the stats package. However, as the reviewer mentions, this example was not included in the original version of the manuscript, where only examples with simulated data were shown. As the reviewer suggests, we have incorporated a new section in the paper where the functions 'kruskal.test' and 'CT_Hypothesis_Test' are compared through the Hollander & Wolfe (1973) dataset and the results obtained are discussed.

Some smaller corrections and/or typos in the text are listed below:

(1) In the abstract the phrase "and a test-based to measure whether" seems to be missing a word after test-based.

We thank the reviewer for noticing this. We have rewritten that as "and a test to measure whether".

(2) The authors write that:

"non-parametric tests is also adequate when there are not enough observations available or when we are analyzing ordinal or nominal data."

Why would a non-parametric test be used when there is limited data? The problem with NP tests is that they lack power but have less strict assumptions. Wouldn't a small dataset generally require a parametric test rather than the other way around? Also, I'm not sure where the "nominal" data part is coming from. None of the tests described in this paper work with purely nominal data.

Parametric tests are more powerful than non-parametric tests and, when only limited data is available, allow to draw stronger conclusions when the necessary hypotheses are met. On the contrary, non-parametric tests do not have this limitation and they offer a good alternative in these cases.

As the reviewer points out, none of the tests described in the paper work with purely nominal data. These could be used employing an encoding to transform nominal into ordinal data. However, to avoid confusion, all references to nominal data have been removed in the manuscript. In particular, we have rewritten the second paragraph of the introduction where the sentence noted above by the reviewer appears.

(3) On the bottom on page 2, there is another mention that the test can

"be used with nominal or ordinal data with nominal"

But then the next paragraph mentions that

"the data can be non-numerical observations if they can be ordered"

Isn't that the definition of ordinal data?

The reviewer is right. Those phrases were confusing. As mentioned before, all references to nominal data have been removed from the manuscript. Particularly, we have rewritten the paragraphs mentioned by the reviewer which now read as follows:

"The Kendall rank correlation coefficient is used as a statistical test to determine whether there is a relationship or dependence between two random variables. The main advantages of this coefficient are: the data can be non-numerical observations if they can be ordered, it is easy to calculate, and the associated statistical test does not assume a known distribution of the population from which the samples are taken."

(4) On page 3 we have

"In our example, the average *range* of the test ... while the average *range* of each"

I think "range" should be "rank", correct?

We thank the reviewer for noticing this. We have made the correction.

(5) On page 9, the definition has "de Concordance" which I think should be "the Concordance".

The reviewer is of course right. We have made the correction.

Package

The package was easy to run and install. It has a small set of functions that are all well-documented. The package uses camel case which is not particularly common in most R packages, but it is at least fairly consistent throughout the main function APIs. As I mentioned above in the article section, it would be nice to include an example in the code that uses a real dataset, similar again to the way the 'kruskal.test' example uses a referenced example.

The authors thank the reviewer for the comments. As said before, an example with the Hollander & Wolfe (1973) dataset (specifically the same example used in the function 'kruskal.test') was already included in the ConcordanceTest package and now it also appears in the new section of the manuscript.

The one concerning major thing about the package is that I am not getting the same results for the Kruskal Test in the 'CT_Hypothesis_Test' and 'kruskal.test' functions. Here is my code:

```
set.seed(12)
A <- c(12,13,15,20,23,28,30,32,40,48)
B <- c(29,31,49,52,54)
C <- c(24,26,44)
Sample_List <- list(A, B, C)

CT_Hypothesis_Test(Sample_List, Num_Sim = 25000, H = 1)
kruskal.test(Sample_List)
```

The test statistics are the same but the p-value in the author's package is slightly lower. Why is this? Is the R package using a conservative analytic approach? this should definitely be exampled in the package and in the paper, perhaps in the section I suggested above comparing these two functions.

We thank the reviewer for pointing this out. In the example shown in the previous code, the p-value for the Kruskal-Wallis statistic in the ConcordanceTest package is lower than that provided by the 'kruskal.test' function. The rationale behind this is that the 'kruskal.test' function uses the χ^2 distribution to approximate the p-value, and this approximation can be quite different from the exact probability distribution if the size of some group is small (less than 5). A comparison between the p-values obtained by simulation in the ConcordanceTest package and the exact ones is made on page 10:

"Note that the approximated p-values obtained by simulation are close to the exact ones, 0.04927 and 0.05223 for the Concordance coefficient and the Kruskal-Wallis statistic, respectively."

Additionally, in the new section of the manuscript, using the example from the Hollander & Wolfe (1973) dataset, we compare our 'CT_Hypothesis_Test' function with the 'kruskal.test' function and discuss why the test statistics with both functions are the same but the p-values differ.

As a minor comment, the progress bar in 'CT_Hypothesis_Test' is helpful to have but would be nice to include a 'verbose' option to turn it off within a script.

As the reviewer suggests, we have added the 'verbose' option that allows the progress bar to be turned off in 'CT_Hypothesis_Test' and we have done the same in the 'CT_Distribution' and 'CT_Critical_Values' functions.