

Dear Heather,

Surely most biological researchers who perform an experiment in which he/she has to compare treatments, wishes them to fall into distinct homogeneous groups. Any statistical procedure used to find those groups, if they exist, will never give biological meaning to them rather than those found by the many statistical measures used to find them. It is up to the biologist to find the biological meaning in the statistical result.

If the F test gives a significant result, which method will give a better partition; Tukey Test, Duncan's Test, Student-Newman-Keuls, Student t-test? Will any one of them give a biological meaning to their results?

We cited two studies published in Brazil showing that simulation results give a better performance to the ScottKnott procedure as compared to the others cited above. We also explained in the article, that the significant levels used in the procedure are significant levels only to each one of the chi-squared test performed by the algorithm and not overall level and that is why the ScottKnott procedure is an **Exploratory Data Analysis** procedure, nevertheless a very good one. I cannot understand which kind of motivating example you had in mind. What does it mean by "*be motivated by a real data example where there is good reason to expect that the treatments divide into distinct groups where the means are equal within*" the example we gave is one appropriate for comparing means, that is, the typical data set that researchers willing to compare means use and that is what the ScottKnott is made for; try to find groups of homogeneous means using a statistical criterion. There is no motivational reason in a real data set. The data set is a text book one, I presume that there is no more motivation than a text book data set.

We could also repeat your last sentence "Even if the ... it doesn't mean it is any more biologically meaningful", using any of the tests named above instead of ScottKnott and the meaning will suit them as well as it suits the ScottKnott.

Finally, I would like to know, what is "the wider R community" which cannot be recommended to use the SK procedure. As we have mentioned the SK procedure is widely used in Brazil by many researchers who happen to be R users, do they belong or not to the wider R community?

Moreover, we read many articles published in the R Journal and by no means we find this article to be of lower level. Otherwise, we would have never sent the article to the R Journal. On the other hand you implied in your answer that the SK

procedure is not suited to be a statistical procedure but your wording does not prove it. The reason is that the procedure is a good one well based and justified and that is why it gives good results and also, no less important, it is simple to use and the results are easily to understand.

We accepted some of criticism made by referees and changed the article to follow the suggestions but now, in this case, we do not agree with you.

Best regards

Enio Jelihovschi

José Claudio Faria

Copy of the e-mails received from Heather Turner

Dear Enio,

Thank you for your resubmission. However after reviewing the revised article and the referee's original comments, I have decided not to forward the article for re-review. The motivating example that you have added only serves to highlight the fact that while biologists may wish their treatment means to fall into distinct homogeneous groups there is often no reason to suspect this to be the case. The method described promotes a false security in statistical significance – if the overall F test for a difference in treatment means is significant, there will typically be many ways to partition the treatments into significantly different "high" and "low" groups. Even if the Scott-Knott algorithm could guarantee to identify the "most significant" grouping factor (which it doesn't) it doesn't mean it is any more biologically meaningful. Therefore I do not think this is an approach to recommend to the wider R community.

Best regards,

Heather

Thank you for your submission to The R Journal. We have now received reports from two referees, see attached.

Both referees identify fundamental problems with using the proposed approach as a multiple comparisons procedure and therefore the article is unacceptable in its current form.

It is possible that the article could be revised to introduce the package in a clustering context as an exploratory tool. However the method would have to be motivated by a real data example where there is good reason to expect that the treatments divide into distinct groups where the means are equal within. The outcome would need to be not immediately obvious from looking at a plot of the data/means. Also appropriate reference would need to be made to other clustering methods available in R.