Principles of Statistical Data Analysis

Homework 3

Assignment

The goals of this homework are:

1. to implement a permutation test yourself.
2. to evaluate this test and compare it to other tests by studying the empirical type I and II error in a Monte-Carlo simulation study.
3. to graphically summarize the performance.

When comparing two independent samples, typically the t-test or the Wilcoxon–Mann–Whitney test is chosen. The t-test has the advantage that the alternative hypothesis is expressed in terms of the mean (a summary measure that is often used) but has the disadvantage that it is sensitive to outliers. The Wilcoxon–Mann–Whitney test, on the other hand, is robust to outliers, but has an alternative hypothesis expressed in terms of the probabilistic index (a summary measure that is less frequently used by practitioners). Because the median is a summary measure that is often used and because it is robust to outliers, it might be valuable to construct a hypothesis test that takes the difference in medians as an effect size.

1. Make an R function `median.test(x,y)` that computes a permutation p-value associated with $H_0 : F_x = F_y$ versus $H_A : \text{median}_x \neq \text{median}_y$.
2. Compare this new test with the permutation t-test and the Wilcoxon–Mann–Whitney test via simulations. You should follow the same set-up as the one discussed on pages 37-38 in the course notes on Permutation and Rank tests.
3. Make a graph similar to the one on page 38 displaying the power for the different tests, sample sizes and distributions.
4. Make a graph similar to the one on page 38, but now under H_0 to compute the empirical type I error.

5. Formulate your main conclusions and recommendations on when to use the median test in practice.

Practicalities

You work together in groups of **3 to 4**. Therefore, agree on how you will organize this before starting the homework. Please detail who did what question. However, note that all group members are responsible for the entire homework, and that you are thus expected to agree with the solutions that your colleague provides. It is expected that the different groups work independently; failure to comply to this, may be interpreted as fraud.

You are expected to submit your written report via the dropbox for HW3 on Ufora. Only one PDF-file can be submitted (Word doc or docx files will not be accepted) and the name of the file should be the family names of the people in the group, ending with HW3, e.g. **DeNevePlevoets-HW3.pdf**. The submission deadline is **Wed, 13 Nov 2019, 23:59**.

Your report should be organized according the different tasks listed above.

- You should include the code of `median.test(x,y)` so that we can read it (evidently, you should clean up this code, make it concise and readable).
- For the Monte-Carlo simulation, you should the include the code of one simulation set up (similar as the code on page 37).
- Include 4 graphs ($n = 20$ or $n = 40$ and simulations under H_A and H_0).
- Formulate your main conclusions and recommendations.
- Provide information on who did what.

In an Appendix to your report you should also include your **R script** with cleaned-up and readable R code!