WEB-BASED APPLICATION PROVIDING GUIDANCE FOR THE CORRECT STATISTICAL TEST SELECTING

Lubomír Štěpánek, Čestmír Štuka, Martin Vejražka
First Faculty of Medicine, Charles University; Faculty of Biomedical Engineering, Czech Technical University in Prague

Keywords: *statistical computations, R, biostatistics, ShinyApps, statistical support, decision making, flowchart*

The problem of choosing the right statistical method in order to analyse our data properly is still the crucial part of a research publication preparation, and – what is more –, the correctness of the choice significantly affects whether the publication will be accepted or rejected by a research journal in the end. We tried to simplify this process in terms of working out an application that gives a helping hand to a researcher handling with statistics in her or his analysis. The application is based on a flowchart diagram navigating to applicable statistical method.

The application is online and free-available. The core of the application is written in language R. The R is a free-as-in-beer and free-as-in-speech programming language and environment for statistical computing and graphics and is widely used among statisticians, econometricians, or biologists – it ensures reliability of computations performed by the application. Code chunks written offline in R language were uploaded online using shiny package to a server dedicated to R calculations. Shiny package is a library written also in R which provides an online framework for R scripts. The application is available at R server of the First Faculty of Medicine, http://shiny.statest.cz. Application uses Czech as a language of choice since we aim on Czech-speaking researchers and students. The flowchart integrated into application helps to make correct statistical decisions depending on the type of inputted data and intended statistical hypothesis; its design arises both from common good practise and from author's experience in this field, too. Statistical methods which are processed so far cover both parametric ones such as t-tests, F-tests or family of ANOVA and others; and non-parametric ones such as Wilcoxon tests, Kruskal-Wallis tests or Friedman ANOVA and others. These methods are directly linked with the flowchart leaf nodes.

The key component of the application is the mentioned flowchart diagram (Fig. 1, in Czech). It enables a researcher who has collected some data and formulated a research hypothesis, to pick the most appropriate statistical method by means of going through the flowchart and step-by-step answering questions. When the researcher reaches an advice which method to choose, a link is provided that allows to perform the suggested method. There is an interface for uploading researcher's data and performing the whole analysis. The key decision points of the flowchart require answering the questions dealing with data types and structures, empirical distributions of data, number of compared samples and whether analysis should compare measures of average or measures of variability.

The project is still in progress and could be interactively updated in response to feedback of its users. In order to reveal our future plans, a way of semi-automated recommendations of which statistical method to use, based on inputted data of a user which could be (pre)processed but non-coerced, suggests itself.

There is no doubt that quality of statistical results presented in any research article significantly affects the overall impact of the publication. The first key step of every statistical analysis is

the choice of an appropriate statistical model. This step is however tricky and difficult for many researchers. A user-friendly tool that makes this choice easier was developed. The direct connection between the best possible statistical method and graphical interface performing computations could improve the user experience.