



WEB-BASED APPLICATIONS FOR WORKING OUT STATISTICAL ANALYSES REQUIRED IN BACHELOR OR MASTER DEGREE THESIS

Štěpánek Lubomír, Feberová Jitka, Dostálová Tatjana

Second Faculty of Medicine, Charles University in Prague

D2.2 MEDICAL INFORMATICS AND INFORMATION SCIENCE

Keywords: *statistical computation application, R, biostatistics, ShinyApps, statistical support*

In order to complete their thesis, bachelor's or master's degree students finishing their studies usually call for any help with statistical analysis of their data. We tried to facilitate this process in terms of working out a system of web-based applications that enable students to input their data into the web-based form of the applications and then download statistical results and plots computed by the applications. Offline texts providing guidance through correct statistical approach and suggesting an optimal final choice of a statistical method are an integral part of the whole concept and system of the applications.

The system consists of a number of online free-available applications. The core of the functionality of the system is based on programming code written in language R. The R is a free software language and environment for statistical computing and graphics and is widely used among statisticians — it ensures reliability of computations performed by the applications, written in R. Code snippets of each application were written offline in R environment and then uploaded online using ShinyApps open-source package, also written in R. Final version of each application is available on www.shinyapps.io server. Interface of each application is user-friendly; Fig. 1 shows interface of application for Harris-Benedict equation (although this is not a statistical topic). A language of choice used by the applications is English because students should be able to understand statistical outputs in foreign language. We work on Czech version of the applications, too. Statistical methods we focus on and provide them for students via the applications cover at the time partly parametric tests of inference such as t-tests, F-tests or family of ANOVA methods and others; partly non-parametric methods such as Wilcoxon tests, Kruskal-Wallis tests or Friedman ANOVA and others.

The project is still in progress and always could be improved or updated. So far we have originated applications dealing with parametric and non-parametric tests as mentioned above. Web address, where is an application available on, follows the general form https://lstepanek.shinyapps.io/name_of_application/, where “name_of_application” is a title of a concrete application, e.g. “t_test” etc. An expected way of usage of each application is — after the choice of the optimal statistical method and appropriate application — an inputting of the student's data into online form, setting parameters of the method via checkboxes and radio-buttons and finally downloading results and plots, provided by interface of the application. The idea of giving rise to such a project of statistical computation applications reflects concrete demands and year-after-year wishes of our students. The applications are aimed not only at bachelor's or master degree's students, but at medical doctors, too. They could particularly appreciate the fact that language of application output plots and tables is English. Offline texts providing guidance through correct statistical approach and links to appropriate applications are necessary in order to guarantee at least a standard quality of statistical section impact on other parts of thesis. Fig. 2 shows a piece of flow chart helping to make a correct statistical decision.

There is no doubt about the fact that a quality of statistical results presented in bachelor or master degree thesis or research article significantly affects the overall impact of the publication. A utilization of power of an open-source statistical language R in order to develop web-based applications written in R is partly one of the way how to try to increase the mentioned quality of statistical analysis in publications and partly a way how to make the process of working out the statistical analysis easier.