Reproducible Research in Public Health

Jinseob Kim[1], **Joohon Sung**[1,★]

1. Complex Disease and Genetic Epidemiology Branch, Department of Epidemiology and Institute of Environment and Health,
School of Public Health, Seoul National University
★Contact author: jsung@snu.ac.kr

Keywords: Reprducible research, LaTeX, Public health, knitr

For most health reasearchers who learned and applied the statistical methods properly had spent long time for learning statistics. Addtionally, statistical methods are ever evolving and updating their knowledge and analytic skills will require the most precious resources-the time. In this study, we aimed to construct pipelines of reproducible statistical analysis in health research. The development of pipelines in this study consists of 1) automatic suggestions of a summary table describing the general characteristics of the study, 2) univariate analysis of both explanatory and outcome variables of a study, 3) graphical presentations of summary and univariate analyses, 4) automatic analysis and tabulations of main results based on frequently used analytical methods in the health research area (e.g., multiple regression, logistic regression, survival analysis, multilevel analysis, genome-wide association study(GWAS)). For example, researchers can obtain tables and figures if they select data set and dependent variables of interest, and define the nature of each variables (e.g. continuous, binomial, count), explanatory variables, and group variable (e.g., sex, region, or unit of random effects).Using *R* package **knitr**, LaTeX and **tex4ht** package in LaTeX with various statistical packages in *R*, we developed a automatic words describing the result tables and figures with PDF or open-document format directly[1, 2]. This automated statistical pipeline tools will help individual researcher in health-related or broader arena to help to reduce their analytical burdens, as well as to conduct appropriate statistical analysis much faster and reliable manner.

References

[1] Andrew, A., A. Zvoleff, B. Diggs, C. Pereira, H. Wickham, H. Jeon, J. Arnold, J. Stephens, J. Hester, J. Cheng, J. Keane, J. Allaire, J. Toloe, K. Takahashi, M. Kuhlmann, N. Caballero, N. Salkowski, N. Ross, R. Vaidyanathan, R. Cotton, R. Francois, S. Brouwer, S. de Bernard, T. Wei, T. Lamadon, T. Torsney-Weir, T. Davis, W. Zhu, W. Wu, and Y. Xie (2013). *knitr: A general-purpose package for dynamic report generation in R*. R package version 1.5.

[2] Cliffe, E. (2012). Methods to produce flexible and accessible learning resources in mathematics: overview document.