StatTag and StatWrap for Conducting Collaborative Reproducible Research

Leah J. Welty

(Northwestern University, Chicago)

28.09.2023, 10.00 (c.t.)

Department of Statistics, Ludwigstr. 33, Room 144
and online via Zoom (Link)
(Meeting-ID: 913-2473-4411; Password: StatsCol22)

Practicing reproducible research is important, but increasingly complex as studies involve more data and statistical code, and larger teams. Adopting reproducible research workflows can be especially daunting for research teams with a diverse set of needs, skills, and expectations for software tools. For example, in medical research, most manuscripts are prepared in Microsoft Word, leaving clinicians to copy and paste, or even re-type, statistical estimates into Word documents. In contrast, statisticians may use R Markdown or Jupyter Notebook to generate reports weaving together statistical results with interpretation, but their collaborators may be unwilling to draft manuscripts in these programs. In addition, teams may struggle to communicate and keep track of information such as: Who worked on the analyses, when, and what decisions did they make? Where is the most recent data? What are the code file dependencies and code libraries? This talk will describe two software tools designed to address these problems – StatTag and StatWrap – both of which grew out of the challenges of conducting ollaborative research in an academic health center. StatTag addresses a need to integrate document preparation in Microsoft Word with statistical code and results from R, Stata, SAS, or Python. StatWrap is an assistive, non-invasive discovery and inventory tool to document the evolution of project, combining automatically collected metadata (e.g., statistical packages, code file dependencies), investigator-supplied documentation (e.g., analysis notes, personnel), and source control. Both StatTag and StatWrap are free, open-source software programs designed to promote the conduct of reproducible research, especially for collaborative teams.

Biography:
Leah J. Welty's research interests include the application of biostatistics to psychiatry and environmental research and the development of software tools for reproducible research. She leads the development team for StatTag and is also the lead biostatistician for the Northwestern Juvenile Project, a large-scale longitudinal study of psychiatric disorders and risky behaviors in delinquent youth, as well as NJP: NextGen, a study of the children of the original Northwestern Juvenile Project participants. She directs the Biostatistics Collaboration Center, the core biostatistics resource for non-cancer research at Feinberg School of Medicine.