openstatsware

Who we are and what we build together

Introducing openstatsware

Background

- Formed on 19 August 2022
- Official working group of the American Statistical Association (ASA) Biopharmaceutical section (BIOP)
- Special Interest Group (SIG) of the European Federation of Statisticians in the Pharmaceutical Industry (EFSPI).
- Cross-industry collaboration (64 members from 41 organizations)
- Homepage: openstatsware.org
- We welcome new members to join!

Motivation

- Open-source software increasingly popular in Biostatistics
  - Rapid uptake of novel statistical methods
  - Unprecedented opportunities for collaboration
  - Transparency of methods and implementation
- Variability in software quality
  - No statistical quality assurance on open-source extension package repositories, e.g. CRAN
  - No industry standard for assessing quality of R packages
- **Reliable software for core statistical analyses is paramount**

Our work

Objectives

- **Engineer selected packages** to fill in gaps in the open-source statistical software landscape, and to promote software tools designed by the working group through publications, conference presentations, workshops, and training courses.
- **Develop good SWE practices** for engineering high-quality statistical software and promote their use in the broader Biostatistics community via public training materials.
- **Communicate and collaborate** with other R software initiatives including via the R Consortium.

We complement the various other R and open source initiatives and statistics SIGs as a bridge between statistical methodology and software. Other groups that we have connections to are Pharmaverse, R Submission Working Group, R Repository Working Group, PSI AIMS, CAMIS, and R Validation Hub.

Workstreams in Package Development

Members from different companies have collaborated on a number of statistical software projects:

- Mixed Models for Repeated Measures (MMRM)
  - Developed the `mmrm` R package for frequentist inference in MMRM
- Bayesian MMRM
  - Developed the `brms.mmrm` R package for Bayesian inference in MMRM
- Health Technology Assessment (HTA)
  - Developed the `maicplus` R package for matching-adjusted indirect comparison (MAIC)
- Bayesian Safety Signal Detection
  - Developed the `SafetySignalDetection.jl` Julia package

Best Practices Dissemination

Our members are widely engaged with teaching and outreach to encourage best practice in statistical software development.

Workshops

- Workshop "Good Software Engineering Practice for R Packages" on world tour
- To teach hands-on skills and tools to engineer reliable R packages
  - Topics: R package structure, engineering workflow, ensuring quality, version control, collaboration and publication, and shiny development
- 5 events in 2023 in Basel, Shanghai, San José, Rockville, and Montreal
- 4 events in 2024 in Zurich, Salzburg, Beijing, and online at R/Pharma APAC

openstatsguide

- Found online here
- Small and concise set of recommendations for package developers
- Opinionated, but aims to be based on experienced majority opinions
- Focus are developers, while users might find complementary "validation" frameworks valuable
- Primarily for statistical packages (not plotting, data wrangling, etc.)
- Generic principles which can be used across functional data science languages R, Python, and Julia
- Concrete tools are mentioned as examples