

Code as Manuscript: Analytical Practices in SAS for Data Efficiency and Reproducibility

Course Syllabus

Instructors

Luke Johnston, MSc, PhD (C)
luke.johnston@mail.utoronto.ca

Daiva Nielsen, PhD
daiva.nielsen@mail.utoronto.ca

Class day, time and location: 2 hours per class is requested

Course description: The purpose of this course is to introduce students to the concept of research reproducibility and present methods of conducting statistical analyses in a more efficient, less error prone way. Tips for increasing the efficiency of collaborative work will also be presented. The content will be tailored specifically to SAS users. The ultimate goal of the course is to demonstrate how to reduce the number of steps required to produce a manuscript from the data analysis stage. Students will be expected to bring their own fully-charged laptop to each class with the appropriate software installed (Git, a text editor and SAS). Emphasis will be placed on hands-on practice activities integrated into each weekly class.

Suggested preparation: Working knowledge of SAS, some experience with the command line or familiarity with computing.

Intended audience: Graduate students or post-doctoral fellows whose primary research is data analysis heavy.

Course objectives: At the end of the course, students will be able to:

- Track their research files with Git to control multiple versions of files.
- Operate GitHub to maintain a back-up of files and have an outlet for efficient collaboration.
- Create macros to condense statistical code writing.
- Use ODS to output results into a more useable form.

Course Schedule

Week 1: Introduction to version control with Git

Week 2: GitHub

Week 3: Macro-coding

Week 4: Output Delivery System (ODS)

Week 5: Review and practice (optional)

Course Resource List

Wilson G, Aruliah DA, Brown CT, Chue Hong NP, Davis M, et al. (2014). Best practices for scientific computing. PLoS Biol. 12(1):e1001745.

Sandve GK, Nekrutenko A, Taylor J, Hovig E (2013) Ten Simple Rules for Reproducible Computational Research. PLoS Comput Biol 9(10): e1003285.

Alsheikh-Ali, Qureshi W, Al-Mallah MH, Ioannidis JP. (2013). Public availability of published research data in high-impact journals. PLoS One. 6(9):e24357.

Laine C, Goodman SN, Griswold ME, Sox HC. (2007). Reproducible research: moving toward research the public can really trust. Ann Intern Med. 146(6):450-3.

Peng RD, Dominici F, Zeger SL (2006). Reproducible epidemiologic research. Am J Epidemiol. 163(9):783-789.

<http://reproducibleresearch.net/> (*an informative blog*)

<http://blogs.plos.org/everyone/2012/08/14/plos-one-launches-reproducibility-initiative/> (*Announcement of PLOS ONE 2012 Reproducibility Initiative*)