

EEES Department Seminar

Fundamentals of Reproducible Research

Dr. Thomas LaBone

MJW Corporation

Aiken, SC

ABSTRACT - Research is defined here to be the process where we: (i) ask a question, (ii) acquire data that we hope can answer the question, (iii) analyze the data, and (iv) draw conclusions from the analysis that are generally applicable to similar situations and data not yet observed.

Research can be high stakes -- a clinical trial for a new cancer treatment for example. Or, it can be mundane, like trying to decide if your analytical instrument is operating properly. The gold standard for demonstrating that the conclusions you reach at the end of your research are valid is replication. Research is replicated when another person independently acquires another dataset, reanalyzes it, and arrives at "more or less" the same conclusions. Replication is not always feasible because it can be expensive, time consuming, unethical, or impossible. A lesser standard is reproduction. Research is reproduced when another person can recreate all the numbers and graphs in your report given your data, code, and associated documentation. There is a bit of a crisis in modern research because an unreasonable amount of published research can't be replicated or reproduced. Failure to replicate someone's work is called science. Failure to reproduce someone's work is more troubling because at first glance one might think this should be easy to do. But, who here has never experienced the situation where a plot in a report can't be reproduced by the author (much less someone else) at a later date? One can't help but to be suspicious of any research that can't be reproduced.

The idea of reproducible research centers around configuring the workflow in your research to make it possible for someone else to readily reproduce all the numerical results and graphs in your report, starting with the original data and documentation on how you manipulated the data. Today we are going to discuss some details of reproducible research.

The software tools used in your research have a huge impact on the effort involved with creating reproducible research and hence on the chances of your work being reproducible. The ubiquitous Microsoft Word/Excel applications do not easily lend themselves to the production of reproducible research, but there are other software packages that do. We will review some freely available applications like the statistical programming language R and the word-processing/typesetting software LaTeX that make this task easier. The goal of this software review is not necessarily to convert you to using these tools, but to illustrate what you should be trying to do with Microsoft Word/Excel if you use them to do your research.

About Dr. LaBone:

Dr. Thomas LaBone is a Health Physicist/Statistician at MJW Corporation, Aiken SC office. Dr. LaBone provides health physics and applied statistical services to government and industrial clients like national laboratories, radiopharmaceutical companies, and companies involved in the commercial reactor fuel cycle. His primary duties are performing internal dosimetry technical calculations (especially for the actinides) and statistics consulting for the NIOSH EEOICPA program. He received his Ph.D. in Biostatistics from the University of South Carolina, Master of Industrial Statistics from the University of South Carolina, MS Radiological Sciences from Rutgers University, and BS Chemistry and Biology from The College of New Jersey.



2:30 PM

Friday, January 26, 2024

Brackett Hall 100

Attendance is mandatory for graduate students enrolled in

EES 8610, EES 9610, and GEOL 8610.

Refreshments following seminar.