

# Simulation in R – Summer, 2015

**Time:** 6/15/15 – 6/19/15 12:30-3:00 and 3:30-6:00

**Instructor:** Sunthud Pornprasertmanit

**Office Hours:** By appointment

**Primary Office:** IMMAP

**Email:** [sunthud.pornprasertmanit@ttu.edu](mailto:sunthud.pornprasertmanit@ttu.edu)

**Class website:** <http://www.myweb.ttu.edu/spornpra/RSIM.html>

## Required

- Laptop with the latest version of R (R 3.2.0 or higher) and Miktex (2.9.0 or higher). If you use Windows, Rtools should be installed (available in CRAN website). Also, R GUI in Windows is horrible so RStudio or NppToR is recommended.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual review of psychology*, 60, 549-576.
- Howard, W. J., Rhemtulla, M., & Little, T. D. (2015). Using Principal Components as Auxiliary Variables in Missing Data Estimation. *Multivariate Behavioral Research*. doi: 10.1080/00273171.2014.999267

## Prerequisites

- Comfortable knowledge of basic statistics that is covered in the first undergraduate statistics class
- Comfortable knowledge of R. Students should be able to run basic statistics in R, such as *t*-test or multiple regression.

## Primary Course Goals and Objectives

1. Acquire a basic understanding of Monte Carlo simulation
2. Develop a skill of assembling R functions to run a simple Monte Carlo simulation
3. Gain knowledge of using programming techniques in R, such as if statement, for statement, and debugging tools
4. Gain knowledge of using efficient programming in R, such as the apply family
5. Gain knowledge of communicating R with other programs to run a simulation
6. Gain knowledge of using R to create figures or tables for publication
7. Gain knowledge of using high-performance computing to run a large-scale simulation

## Expectations

I do not expect students to master all programming techniques within the class. This class is only a gentle introduction for simulation in R. Thus, I expect students to have basic knowledge in simulation in R. Students should develop the knowledge and skills after the class to be the master in one day.

Because this class is small, I do expect you to ask questions. All students and I can wait for everyone to go together. If you do not understand anything, please speak up. I make plenty of time available for questions so you will not make the class behind the schedule.

I expect students to be able to run a small simulation study after the class. Thus, I expect you to talk with your advisor or IMMAP personnel to come up with a project that both you and IMMAP are interested in. I hope that the project that you run in this class will be a baby step for a good publication in the near future.

## Evaluation

*In-class exercises (40%).* I will assign one assignment at the end of each day. Each assignment is a small simulation. Students should write a brief report that contains (a) the goal of this simulation, (b) the simulation method, (c) the result of the simulation (with figures or tables), and (d) the discussion of the results. I am okay if each section only has one or two paragraphs. R script is attached as an appendix. Comments should be provided in each group of lines (says 5-15 lines) to tell me the goal of the chunk of codes. The due time of each assignment is one day after the assignment date by 10 am. If the assignments need to be redone, all fixed assignments must be sent to me by 6/26.

*Simulation project (40%).* You will write a manuscript detailing your simulation from a project conducted in conjunction with your adviser. The simulation should be small so that the running time is no longer than one day in HPC. The manuscript should be similar to publication in methodology journals, such as Educational and Psychological Measurement. At least one figure or one table is required for showing your simulation results. The due date is 7/31 by midnight via email. R script should be sent via email in separate files as well. If you have multiple files, please explain each file in the email. I will be around before the due date. Please contact me via email or in person if you need any assistance.

*Simulation outline (10%).* You will write the ideas that you are using for the simulation project. The outline should cover the importance of your research questions and the short description of how would you implement the simulation. The length should not be over 500 words. The due date is 6/26 by midnight.

*Simulation proposal (10%).* You will write the proposal that covers the introduction and method sections of your simulation project. The details should be enough such that other quantitative experts can run your simulation based on information provided in the proposal. The R script of the main analysis function in your simulation should be provided too. The due date is 6/29 by midnight via email.

## Class Schedule

<i>Date</i>	<i>Topic</i>
June 15 (1)	Getting started
June 15 (2)	Figures, functions
June 16 (1)	If statement and the <code>apply</code> function
June 16 (2)	Communication with external programs using R
June 17 (1)	<code>sapply</code> , <code>lapply</code> , <code>mapply</code> functions and parallel processing in R
June 17 (2)	Data generation, missing data, and debugging
June 18 (1)	High performance computing, pseudo random number generator
June 18 (2)	Making a package, documentation (if time allows)
June 19 (1)	Latex basics, Sweave (if time allows)
June 19 (2)	Object-oriented programming (if time allows)
<b>June 26 (12 am)</b>	<b>Due time: Simulation proposal</b>
<b>June 29 (12 am)</b>	<b>Due time: Simulation outline</b>
<b>July 31 (12 am)</b>	<b>Due time: Simulation project</b>