

INTRODUCTION TO R STATISTICAL ANALYSIS SOFTWARE

Summer 2022
08/15/22 – 08/19/22

Instructor:	SongYi Paik	Time:	1:00PM – 4:00PM
Email:	songyi@umn.edu	Place:	Ruttan Hall 135B or Zoom
Office Hours:	by appointment		(A Zoom link is available on Canvas)

Objectives

This course is designed as an introduction to R statistical analysis software for incoming graduate students. Instead of covering every R skill, you will build a solid foundation to prepare you for the more in-depth courses including Econometric Analysis (APEC8211-8212), Programming for Econometrics (APEC8221), and Big Data Methods in Economics (APEC8222). Being familiar with the basics of R, you will be able to handle working with new functions and packages with little trouble.

By the end of this course, you should be able to:

1. Install and configure software necessary for a statistical programming environment
2. Access R packages
3. Read data into R
4. Create vectors and matrices
5. Use dataframe and datatable
6. Write R functions
7. Run a linear regression
8. Make data usable by merging and reshaping
9. Generate random numbers
10. Simulate data using for loop
11. Visualize data

Reading

Norman Matloff, *The Art of R Programming*, No Starch Press, 2011.

Things To Do Before The Course Begins

- You need to install R and R studio, which is an integrated development environment designed for R. Since R is an open-source software, you can download all necessary software and various packages for R at no cost on your laptop and should bring yours to class. **You can find installation instructions on Canvas/Modules/Setup (Install R and Rstudio)/set_up.html.**
- Please submit the short pre-course survey sent to you by August 12th (Friday).

Course Content

- **Time Table:** Each class will be structured as follows:

1. First 10-15 minutes: lecture
2. Second 15-20 minutes: activity (practice + review)
3. Third 10-15 minutes: lecture
4. Fourth 15-20 minutes: activity (practice + review)
5. 5-minute break

Repeat this 60-minute block schedule three times.

- **Topics** (tentative)

1. **Day 1: Vectors and Matrices**

- Intro R
- Vectors
- Matrices
- Lists
- Factors
- Functions
- Logical operators

2. **Day 2: Data Frames & Linear Regression**

- Dataframe
- Datatable
- Packages
- Read data
- Linear regression

3. **Day 3: Managing Data**

- Reshaping
- Merging
- Cleaning
- Save data
- API

4. **Day 4: Iteration**

- Random number generation
- Loop functions
- Simulation
- Parallel processing: foreach, future, doFuture, and doRNG packages

5. **Day 5: Data Visualization**

- Plot function
- ggplot2 package
- Generating tables