

# RStudy 2020: Basic Statistics with R

Sanghoon Park

Fall 2020

E-mail: [sp23@email.sc.edu](mailto:sp23@email.sc.edu)

Office Hours: TH 1-3 p.m. (US)

Office: #305 Gambrell Hall

Web: [shpark.netlify.app/](http://shpark.netlify.app/)

Class Hours: M 8-9 a.m. (US) 9-10 p.m. (ROK)

Class Room: [Zoom online](#) (From Sep. 14th 2020)

---

## Study Description

이 R 스터디는 선형회귀분석을 가지고 변수들 간의 관계를 측정 및 해석하는 통계기법을 살펴보는 것을 목적으로 한다. 구체적으로 R에 관한 기본적인 작동 원리 및 코딩을 위한 팁 등을 함께 살펴보고, 나아가 통계학적 기본 지식들(확률, 분포, 추론 등)에 대한 리뷰 및 유의성 검정, 최소자승법(ordinary least squares, OLS)을 이용한 선형회귀모형 추정에 관한 내용들을 개략적으로 훑어볼 것이다. 만약 시간이 허락한다면, 비연속적 종속변수를 추정하기 위한 일반화 선형모형(generalized linear models, GLM)과 어떤 통계 모형을 사용할 것인지에 관한 선택의 문제(모델 진단, 모델 선택 과정, 결측치 문제)도 함께 다루어 볼 것이다. 이 스터디는 단순히 통계학적 지식을 이론적으로 학습하기보다는 재생산 가능한 자료를 이용하여 R을 이용한 통계분석 및 프로그래밍을 실습하고, 나아가 그 분석 결과를 실제 활용할 수 있는 방식으로 보여줄 수 있는 능력을 함양하는 데 목적이 있다. 스터디를 마치게 되었을 때, 구성원들은 선형 회귀를 이용하여 경험적 관계를 평가하고, 이후 보다 고급 통계 모델 및 기법들을 이해하는 기초를 마련할 수 있게 될 것으로 기대한다.

## Expected Learning Outcomes

스터디를 통해 구성원들은 다음과 같은 성과를 얻을 것으로 기대한다.

- R을 이용해 데이터를 불러오고, 변형하여 통계적 검정을 수행하고, 나아가 선형모형을 추정할 수 있게 될 것이다.
- OLS 접근법의 가정들을 이해하고, OLS를 통해 얻은 결과를 해석할 수 있게 될 것이다.
- 모델 특정(specification), 혹은 모델 특정(misspecification)의 오류와 관련된 문제들을 진단하고, 바로잡을 수 있게 될 것이다.
- 모델 특정과 연구설계(research design)에 관련하여 제기될 수 있는 문제들을 설명할 수 있게 될 것이다.

## Materials for Use

이 스터디는 R을 사용하기 위한 컴퓨터를 필요로 하며, 정기적으로 [Zoom](#) 을 통한 온라인 미팅을 실시한다. 스터디는 다음과 같은 주교재를 중심으로 운영된다:

- Fox Jr., John. 2016. *Applied Regression Analysis and Generalized Linear Models, Third Edition*. Thousand Oaks, CA: SAGE Publications.

R은 프로그래밍 언어이면서 컴퓨팅과 그래픽을 통해 통계적 분석을 할 수 있도록 하는 환경이라고 할 수 있다. R 과 관련된 자료는 다음의 링크를 통해 확인할 수 있다.

- R 다운로드: <https://www.r-project.org/>.
- RStudio는 R이라는 프로그래밍 언어를 좀 더 사용자 친화적인 방식으로 제공하는 인터페이스이다: <https://rstudio.com/products/rstudio/download/>

이외에 통계학 이론에 관해서는 다음의 자료를 참고한다.

- Jeffrey M. Wooldridge. *Introductory Econometrics: A Modern Approach, 5th (or later) edition*. Cengage Learning.
- Andrew Gelman and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel Hierarchical Models*. Cambridge University Press.

R 프로그래밍 및 코딩과 관련된 자료는 아래를 참고할 것을 권한다.

- Hadley Wickham and Garrett Grolemund. *R for Data Science*
- Hadley Wickam. *Advanced R 2nd Edition*
- Hadley Wickam. *ggplot2: Elegant Graphics for Data Analysis*

주마다 참고할 만한 리딩자료는 드랍박스의 Textbook 폴더에, 스터디에 활용할 슬라이드 자료는 Documents (Slides) 폴더에 주차별로 추가될 예정이다.

## Study Schedules

### Week 01, 09/07 - 09/11 : Introduction to R

첫 주는 기본적인 R에 관한 내용과 코딩 스킬 및 에티켓에 관한 내용, 그리고 데이터를 불러들이고 관리하는 것에 대하여 논의한다.

#### *Required*

- R Script for Week 1

#### *Recommended*

- “Coding Style and Good Computing Practices.”

### Week 02, 09/14 - 09/18 : Statistical Models and Social Science (& Tidyverse)

#### *Required*

- Fox, Chapter 1
- R Script for Week 2

## **Week 03, 09/21 - 09/25 : Types of Data; Visualizaing and Transforming Data**

### *Required*

- Fox, Chapter 3 & 4
- R Script for Week 3

### *Recommended*

- “Data Visualization in Sociology.”
- “Data Visualization with R.”

## **Week 03, 09/21 - 09/25 : Probabilities and Distributions I**

### *Required*

- R Script for Week 4

## **Week 04, 09/28 - 10/02 : 추석 연휴**

## **Week 05, 10/05 - 10/09 : Probabilities and Distributions II**

### *Required*

- R Script for Week 5

## **Week 06, 10/12 - 10/16 : Foundation of Inference I**

### *Required*

- Fox, Chapter 1 (re-read) & 2
- R Script for Week 6

### *Recommended*

- “Do Statistical Reporting Standards Affect What Is Published?”

## **Week 07, 10/19 - 10/23 : Foundation of Inference II**

### *Required*

- R Script for Week 7

### *Recommended*

- “Abandon Statistical Significance”
- “The Statistical Crisis in Science.”

## **Week 08, 10/26 - 10/30 : Introduction to Linear Regression I**

### *Required*

- Fox, Chapter 5
- R Script for Week 8

### *Recommended*

- “Is OLS Dead?”

## **Week 09, 11/02 - 11/06 : Introduction to Linear Regression II**

### *Required*

- Fox, Chapter 6
- R Script for Week 9

### *Recommended*

- “Statistical Models and Shoe Leather”

## **Week 10, 11/09 - 11/13: Simple Linear Regression**

### *Required*

- R Script for Week 10

### *Recommended*

- “Let’s Put Garbage-Can Regressions and Garbage-Can Probits Where They Belong.”
- “Strategic Misspecification in Regression Models.”

## **Week 11, 11/16 - 11/20: Multivariate Linear Regression I**

### *Required*

- Fox and Weisberg, Chapter 4
- R Script for Week 11

### *Recommended*

## **Week 12, 11/23 - 11/27: Multivariate Linear Regression II**

### *Required*

- Fox and Weisberg, Chapter 5
- R Script for Week 12

### *Recommended*

## Week 13, 11/30 - 12/04: Understanding of Interactions

### *Required*

- R Script for Week 13

### *Recommended*

- “Understanding Interaction Models: Improving Empirical Analyses.”
- “How Much Should We Trust Estimates from Multiplicative Interaction Models? Simple Tools to Improve Empirical Practice.”