

2015 Division of Public Health Sciences R Primer

July 6-8, in the Doll & Hill Room, Taylor Avenue Building.

1-credit course for graduate students, postdoctoral fellows, and residents.

- Instructional Staff:
 - Jeff Gill, Instructor/Course Master, [email](#)
 - Jung Ae Lee, Instructor, [email](#)

- Description: This is a short 1-credit primer to introduce the R Statistical Environment to new users. R is "a freely available language and environment for statistical computing and graphics which provides a wide variety of statistical and graphical techniques: linear and nonlinear modelling, statistical tests, time series analysis, classification, clustering, etc. The goal is to give you a set of tools to perform sophisticated statistical analysis in medicine, biology, or epidemiology.

- Competencies: At the conclusion of this primer participants will: be able to manipulate and analyze data, write basic models, understand the R environment for using packages, and create standard or customized graphics.

- Prerequisite Details: This primer assumes a knowledge of basic statistics as taught in a first semester undergraduate or graduate sequence. Topics should include: probability, cross-tabulation, basic statistical summaries, and linear regression in either scalar or matrix form.

- Grading:
 - Attendance/Participation: 20%
 - Data Assignment 1: 40%
 - Data Assignment 2: 40%

On the first and second day a data assignment will be given. It is due before the commencement of the next class session. Email your work to the instructor for that assignment.

- Datasets and Assignments:
 - Assignment 1:
 1. Read the description and download the [Hemodialysis Data](#)
 2. Describe each variable numerically and graphically.
 3. Find an interesting relationship between two dichotomous variables and describe it with a 2 X 2 table.
 - [Colon Cancer](#) for Tuesday lab work.
 - Assignment 2:

[Description](#)
[Data.](#)

-
- Available Reading: The following are high quality monographs that are 100% free online.
 - [An Introduction to R](#)
 - [Using R for Data Analysis and Graphics, J.H. Maindonald](#)
 - [simpleR - Using R for Introductory Statistics, John Verzani](#)

- [The R Guide](#)
- [Analysis of Epidemiological Data Using R and Epicalc, Virasakdi Chongsuvivatwong](#)
- [Statistics Using R with Biological Examples, Kim Seefeld and Ernst Linder](#)
- [An Introduction to R: Software for Statistical Modeling & Computing, P. Kuhnert & B. Venables](#)

There are also *many* commercial books for sale; ask one of the instructors for suggestions in your area of interest.

- Course Outline:
 - Module 1, Monday, July 6, 2015, 8:30-12, [Jeff Gill](#)
 1. Downloading and Installing R
 2. Downloading and Installing R Packages
 3. Setting Up Help
 4. Basic Syntax (comments, naming conventions etc)
 5. Data Types and Data Structures
 6. Basic Operations
 7. Data Import and Export
 8. Quitting R and Saving R Objects
 9. Basic Tabular Analysis
 10. [slides for introduction](#) (do not print).
 - Module 2, Monday, July 6, 2015, 1:30-5, [Jeff Gill](#)
 1. Basic Plotting Commands
 2. Plotting Categorical Data
 3. Basic Two-Dimensional Plotting Commands
 4. Combining Different Graphs
 5. Important Ways To Export Your Graphs
 6. Setting Up The Graphics Window
 7. Following Trends
 8. Illustrating The Law of Large Numbers
 9. Illustrating The Central Limit Theorem
 10. ROC Curves
 11. Graphing Networks
 12. [slides for plotting](#) (do not print).
 - Module 3, Tuesday, July 7, 2015, 8:30-12, Jung Ae Lee.
 1. Data
 2. Univariate Data
 3. Bivariate Data
 4. Regression Analysis
 5. [slides for all topics](#).
 6. [Slides Day 2](#).
 - Module 4, Tuesday, July 15, 2014, 1:30-5, Jung Ae Lee.
 1. Multivariate Data
 2. Multiple Linear Regression
 3. [Slides Day 3](#).
 - Module 5, Wednesday, July 8, 2014, 8:30-12, [Jeff Gill](#)
 1. Defining Functions
 2. Multiple Arguments
 3. Naming Your Functions
 4. Loops In Functions
 5. Termination
 6. Counting Rules and Permutations

- Module 6, Wednesday, July 8, 2014, 1:30-5, [Jeff Gill](#)
 1. Sampling
 2. Generating Samples in R
 3. Random Imputation
 4. Monte Carlo Introduction
 5. Basic Monte Carlo Integration
 6. Rejection Sampling
 7. Bootstrapping for Standard Errors
 8. [slides for functions](#) (do not print).
 9. [slides for sampling](#) (do not print).

- Helpful Websites:
 - [The Comprehensive R Archive Network](#)
 - [Washington University Guide to R Resources](#)
 - [R Biostat Resources at Vanderbilt](#)
 - [An R Reference Card](#)