

# BIOST 2094 - Statistical Computing in R

## Spring 2015

Time and Place: Tuesday 10:00-11:55 AM, WWPH 1201

Instructor: Tianzhou Ma (Charles), MS  
E-mail: tim28@pitt.edu

TA: Kidane Ghebrehawariat  
E-mail: kig11@pitt.edu  
Office Hours: Thursday 11:00 AM - 12:55 PM, Crabtree A443 (inform TA by email at least one day before the office hour)

### Course Description

BIOST 2094 is an introductory course in statistical computing using R for graduate level biostatistics students. The course is divided into two parts. In the first part of the course, students learn programming techniques for writing R programs that will do statistical analyses not available in standard software packages. The second part of the course covers how to perform routine statistical methods in R.

Prerequisites: BIOST 2041, 2042

### Learning Objectives

1. Write reliable and transparent programs
2. Produce informative graphical descriptions of data
3. Import, export, and manipulate datasets
4. Analyze data using descriptive and inferential statistics
5. Analyze data through model fitting

### Textbooks:

There is no required textbook. The following books are recommended references:

- John Chambers, *Software for Data Analysis: Programming with R*, Springer, 2008.
- Michael J. Crawley, *The R Book*, 2nd edition, Wiley, 2012.
- Brian Everitt and Torsten Hothorn, *A Handbook of Statistical Analyses Using R*, Chapman & Hall/CRC, 2006.
- Owen Jones, Robert Maillardet, Andrew Robinson, *Introduction to Scientific Programming and Simulation Using R*, Chapman & Hall/CRC, 2009.

## Websites

- R Website <http://www.r-project.org>
- R Manuals <http://cran.r-project.org/manuals.html>
- Contributed Manuals <http://cran.r-project.org/other-docs.html>
- Patrick Burns, The R Inferno <http://www.burns-stat.com/>
- R Journal <http://journal.r-project.org/>
- R Search engine (Internet) <http://www.rseek.org/>
- R Search engine (R documents) <http://search.r-project.org/nmz.html>
- R-bloggers <http://www.r-bloggers.com>
- Quick-R <http://www.statmethods.net>

## Student Evaluation and Grades

Course grades will be based on a weighted average of,

- Homework assignments 40%
- Project (individual) 30%
- Project (group) 30%

The cut-offs for computing letter grades will be: A, 100%-90%; B, 89%-80%; C, 79%-70%; D, 69%-60%; and F, <60%. Plus-minus grades will be assigned by dividing the respective intervals into thirds. Discussions of homework and project among students are allowed but each student has to write their own solution. Cheating and plagiarism is strictly not allowed and may be reported to the university. See the University of Pittsburgh's Policy on Academic Integrity at <http://www.provost.pitt.edu/info/ai1.html>

## Homework

There will be 4 homework assignments. Students will turn-in a hard copy of their R codes and outputs in-class and submit an electronic copy via Blackboard; R files must be sent to Blackboard before the beginning of class, 10:00am. Late homework will not be accepted. *Use white space and include clear comments to make code readable.*

## Project

The goal of the project is to develop an R package that will be useful to other statisticians and R users. The instructor would randomly assign students to form small groups to work on a topic from a list provided by the instructor. The project will have two parts, an individual and group component. For the individual part each group member will submit a well-defined independently executable function that contributes to the overall purpose of the package and will be graded individually. The group will then define classes and write a set of methods to accompany the functions, write the documentation and finally create an R package of all the materials the group created. Each group will give a class presentation on their package. As part of the presentation the group will distribute their package to the class and teach the class how to use it.

## Lecture Schedule

Date	Due	Topic
January 6, 2015		Lecture 1: Getting Started with R
January 13, 2015		Lecture 2: Data Structures
January 20, 2015	Homework 1 Distributed	Lecture 3: Data Types
January 27, 2015	<b>Homework 1 Due</b>	Lecture 4: Control Structures
February 3, 2015		Lecture 5: Functions
February 10, 2015	Homework 2 Distributed	Lecture 6: Graphics
February 17, 2015	<b>Homework 2 Due</b>	Lecture 7: Exception Handling and Debugging
February 24, 2015		Lecture 8: Object-Oriented Programming
March 3, 2015	<b>Individual Project Due</b>	Lecture 9: Creating an R Package
March 10, 2015	Spring Break	No Class
March 17, 2015		Lecture 10: Numerical Methods
March 24, 2015	Homework 3 Distributed	Lecture 11: Dataframes
March 31, 2015	<b>Homework 3 Due</b>	Lecture 12: Basic Statistical Methods
April 7, 2015	Homework 4 Distributed	Lecture 13: Statistical Models
April 14, 2015	<b>Homework 4 Due, Group Project Due</b>	Group Presentations

## Academic Integrity

Students in this course will be expected to comply with the University of Pittsburgh's Policy on Academic Integrity (<http://www.provost.pitt.edu/info/ai1.html>). Any student suspected of violating this obligation for any reason during the semester will be required to participate in the procedural process, initiated at the instructor level, as outlined in the University Guidelines on Academic Integrity. This may include, but is not limited to, the confiscation of the examination of any individual suspected of violating University Policy. Furthermore, no student may bring any unauthorized materials to an exam, including dictionaries and programmable calculators.

## Disability Services

If you have a disability that requires special testing accommodations or other classroom modifications, you need to notify both the instructor and Disability Resources and Services (<http://www.studentaairs.pitt.edu/drswelcome>) no later than the second week of the term. You may be asked to provide documentation of your disability to determine the appropriateness of accommodations. To notify Disability Resources and Services, call (412) 648-7890 (Voice or TTD) to schedule an appointment. The Disability Resources and Services office is located in 140 William Pitt Union on the Oakland campus.