

**Stat – 331 Statistical Computing with R (4)**

Fall 2015

**1. Catalog Description**

**STAT 331 Statistical Computing with R (4)**

Import, manage, and clean data; use of regular expressions; functional and object-oriented programming; graphical, descriptive, and inferential statistical methods; random number generation; Monte Carlo methods including resampling, randomization, and simulation: 4 lectures. Prerequisites: STAT 252, STAT 302, STAT 312, or STAT 313 and CSC 101, CSC 235, CSC 237, or BUS 290; or permission of instructor.

**2. Required Background and/or Experience**

Prerequisite: IME 326, STAT 252, STAT 302, STAT 312, or STAT 313 and CPE/CSC 101, CPE/CSC 235, or BUS 290 or STAT 330; or permission of instructor. Basic knowledge of statistics and computers.

**3. Expected Outcomes**

The student should be able to:

- a. Import, manage, and clean data from a wide variety of data sources.
- b. Use regular expressions and other functionality for data management.
- c. Understand the object oriented programming environment and basic functionality.
- d. Produce and interpret graphics for exploratory data analysis and report presentations.
- e. Write efficient and well-documented code for descriptive and inferential statistics.
- f. Understand how random numbers are generated from probability models.
- g. Program Monte Carlo experiments and simulations.
- h. Extend their skill set independently through help documentation and online resources.

**4. Suggested texts or References**

John Braun and Duncan Murdoch. *A First Course in Statistical Programming with R*. Cambridge University Press. 2007.

Peter Dalgaard. *Introductory Statistics with R*. Springer. 2008.

Deborah Nolan and Duncan Temple Lange. *Data Science in R: A Case Studies Approach to Computational Reasoning and Problem Solving*. Chapman and Hall/CRC. 2015.

Deepayan Sarkar. *Lattice: Multivariate Data Visualization with R*. Springer. 2008.

Phil Spector. *Data Manipulation with R*. Springer. 2008.

Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer. 2009.

**5. Minimum Student Materials**

USB flash drive or external disk

**6. Minimum University Facilities**

Classroom with PC workstations, R software, chalk/whiteboard and overhead projector. Video projection equipment is preferable.

## 7. Description of Content and Method

TOPICS	NUMBER OF LECTURES
<b>A. INTRODUCTION TO R</b>	2
1. Console, packages, scripts, directories	
2. Mathematical and logical operators	
<b>B. DATA STRUCTURES</b>	3
1. Vectors	
2. Matrices	
3. Arrays	
4. Data Frames	
5. Lists	
<b>C. DATA MANAGEMENT</b>	6
1. Reading data from local files	
2. Reading data via a network connection	
3. Data types	
4. Regular expressions	
5. Cleaning, subsetting, merging	
6. Data summaries	
<b>D. GRAPHICS</b>	5
1. Basic graphics	
2. Higher-level graphics (e.g. <i>ggplot2</i> , <i>lattice</i> )	
<b>E. PROGRAMMING</b>	6
1. Object-oriented programming	
2. Iteration	
3. Execution control	
4. Functions	
<b>F. STATISTICS AND PROBABILITY</b>	5
1. Continuous and discrete distributions	
2. Simulating random variables and statistics	
3. Descriptive statistics	
4. Introductory inference	
5. Regression	
<b>G. MONTE CARLO AND RESAMPLING METHODS</b>	5
1. Simulation	
2. Randomization tests	
3. Bootstrapping	
<b>H. OTHER TOPICS AT INSTRUCTOR DISCRETION</b>	4
1. Cross-validation in statistical models	
2. Smoothing	
3. Other modern statistical methodologies	
	<b>Total</b>
	<b>36</b>

## METHOD

Material will be presented in a lecture format and class discussions. Applied material will be presented using lectures, real-time computer-screen projections, and supervised student computer-activities.

## 8. Method of Evaluating Outcome

Homework assignments, quizzes, projects, and examinations.