**Class Time: Mondays, 6:00-9:30 PM and June 2 (05/22/2017 – 08/07/2017)**
**Class Location: L1110**

**Instructor:** Ching-Ti Liu, Ph.D. Telephone: (617) 638-7752
Office: CT (Crosstown) Room 329
E-mail: ctliu@bu.edu
Office hours: 3:00 ~ 4:00 PM on Thursdays or by appointment.

**Course description:**
The goal of this course is to provide students with the mathematical and practical background required in the field of data analytics. Starting with an introduction to probability and statistics, the R tool is introduced for statistical computing and graphics. Different types of data are investigated along with data summarization techniques and plotting. Data populations using discrete, continuous, and multivariate distributions are explored. Errors during measurements and computations are analyzed in the course. Confidence intervals and hypothesis testing topics are also examined. The concepts covered in the course are demonstrated using R. This is a laboratory course.

**Prerequisite: CS546 (Quantitative Methods for Information Systems) or equivalent background.**
**Credits: 4 credits.**
**Course website: learn.bu.edu**

**Course's rationale, format and objectives**
The goal of this course is to provide a supportive, hands-on environment for students to gain mathematical and practical experience with R. Weekly sessions will include: demonstration of the various statistical computing skills in R and small group and class wide discussion. Each week the instructor will provide a dataset, along with a set of questions, for group discussion. Weekly homework assignments will be required for this class and the homework will emphasize applications of statistical computing to real datasets. For the final project students will work in group using R to comprehensively analyze a dataset and produce an in-class oral presentation and a written report.

At the end of this course students will be able to:
- Acquire the basic concepts of R
- Model random experiments using R
- Analyze univariate, bivariate, and multivariate data
- Learn various discrete and continuous probability distributions
- Understand sampling methods and errors
- Interpret confidence intervals
- Associate Type-I and Type-II errors while doing hypothesis testing

## Text (Recommended but not required):


**Kerns, G. J. (2010).** ***Introduction to probability and statistics using R.***

Publisher: G. Jay Kerns.

ISBN-13: 978-0557249794

ISBN-10: 0557249791

This book is available for download.


**Verzani, J. (2014).** ***Using R for introductory statistics*** **- 2nd Edition.**

New York: Chapman & Hall.

ISBN 9781466590731.

This book can be purchased from Barnes & Noble at Boston University.

**Course Requirements**:

It includes assignments (25%), quiz (10%), midterm exam (20%), final project (20%), final exam (20%) and class participation (5%). No late assignment will be accepted if no explanation for the lateness was given before its due date. Students are encouraged to discuss with their classmates but need to independently submit their homework assignment.

1. Assignments (25% of final grade)
    - **Goal:** Six homework assignments allow students to apply their understanding of the concepts covered during the weekly lecture
    - **Logistics:** (1) Students are encouraged to discuss with each other but need to submit assignments independently. (2) **In course website:** Students should submit one electronic copy (both R code and write-up) through Course website (learn.bu.edu) before the start of class.
    - **Grading:** All homework assignments will be graded and counted into the final grade.

2. Quiz (10% of final grade)
    - **Goal:** Four short quiz enhance students' understanding and allow them to apply their understanding of the concepts within each module covered in several lectures.
    - **Logistics:** (1) Short online quiz help students review the material including statistical concepts and computing skills. (2) Students will work on the quiz independently. Short online quiz helps students develop capability to apply their knowledge independently.
    - **Grading:** All quiz will be graded and counted into the final grade.

3. Final Project (20%)
    - **Goal:** This project serves two purposes. One is for you to utilize the concepts/skills covered in this class on a real data set. Second, it provides a chance for you to practice and discuss the concepts in a group setting so that you can get feedbacks from your group members
    - **Logistics:** (1) Students will work on final project in groups of 2~3 students. (2) Students will submit a proposal on their dataset and analysis plan in group. (3) Students will analyze their own dataset in group. (4) Students will present their final projects in class as group (5) Students will actively participate in other groups' presentation.
    - **Grading**: 10% for in-class **group presentation** and discussion and 20% for **individual written report.** Both parts will be graded based on the accuracy, clarity, and presentation. Over discussion, students are encouraged to actively participate other groups' presentation.

4. Mid-term and Final exam (20% for midterm and 20% for final)
    - **Goal:** This part is to enhance students' understanding and allow them to apply their understanding of the concepts across different modules/lectures
    - **Logistics:** (1) These exams help students review the material including statistical concepts and computing skills. (2) Students will work on the in-class exam independently. These helps students connects different concepts and develop capability to apply their knowledge independently.
    - **Grading:** All exams will be graded and counted into the final grade.

5. Class Participation (5%, for in class participation and course website discussion**)**
    - **Goal:** This part is to prepare you to be an active learner and extend your learning outside of class by stating your questions in a clear and informative way, by addressing other classmates' question and by sharing your learning.
    - **Logistics:** Students are encouraged to help with each other during this learning process. (1) Students will present their either questions or answers to the in-class exercise (2) Students who have mastered the in-class material will help to address the questions raised by their classmates both in class setting and course website discussion forum. (3) Students are also encouraged to post messages on the course website to share the extended materials or programming tricks which are not covered in class.
    - **Grading:** Students are not graded by attendance. However, students will be graded based on whether they volunteer to present their work for in-class exercise, actively participate in the in-class discussion and successfully address others' question or share the trick in the course website.

## Tentative Course schedule Chart

| Class | Date | Topic | Assignment | Quiz |
|---|---|---|---|---|
| 1 | 22-May | Module 1 (Introduction) | | |
| 2 | 29-May | Module 1 and Module 2 (Probability) | | Quiz 1 |
| 3 | 5-Jun | Module 2 (Probability) | HW1 | Quiz 2 |
| 4 | 12-Jun | Module 3 (Data Description) | HW2 | |
| 5 | 19-Jun | Module 3 and Module 4 (Distribution) | | Quiz 3 |
| 6 | 26-Jun | **Mid-term examination** and Module 4 (Distribution) | HW3 | |
| 7 | 3-Jul | Module 4 | | Quiz 4 |
| 8 | 10-Jul | Module 5 (Sampling and Errors) | HW4 | |
| 9 | 17-Jul | Module 5 and Module 6 (Estimation of Performance) | HW5 | |
| 10 | 24-Jul | Module 6: Estimation of Performance | HW6 | |
| 11 | 31-Jul | **Final Examination** | | |
| 12 | 7-Aug | **Student project presentation** | | |

# Academic Conduct Policy

For the full text of the academic conduct code, please go to http://www.bu.edu/met/for-students/met-policies-procedures-resources/academic-conduct-code/.

## A Definition of Plagiarism

"The academic counterpart of the bank embezzler and of the manufacturer who mislabels products is the plagiarist: the student or scholar who leads readers to believe that what they are reading is the original work of the writer when it is not. If it could be assumed that the distinction between plagiarism and honest use of sources is perfectly clear in everyone's mind, there would be no need for the explanation that follows; merely the warning with which this definition concludes would be enough. But it is apparent that sometimes people of goodwill draw the suspicion of guilt upon themselves (and, indeed, are guilty) simply because they are not aware of the illegitimacy of certain kinds of "borrowing" and of the procedures for correct identification of materials other than those gained through independent research and reflection."

"The spectrum is a wide one. At one end there is a word-for-word copying of another's writing without enclosing the copied passage in quotation marks and identifying it in a footnote, both of which are necessary. (This includes, of course, the copying of all or any part of another student's paper.) It hardly seems possible that anyone of college age or more could do that without clear intent to deceive. At the other end there is the almost casual slipping in of a particularly apt term which one has come across in reading and which so aptly expresses one's opinion that one is tempted to make it personal property."

"Between these poles there are degrees and degrees, but they may be roughly placed in two groups. Close to outright and blatant deceit-but more the result, perhaps, of laziness than of bad intent-is the patching together of random jottings made in the course of reading, generally without careful identification of their source, and then woven into the text, so that the result is a mosaic of other people's ideas and words, the writer's sole contribution being the cement to hold the pieces together. Indicative of more effort and, for that reason, somewhat closer to honest, though still dishonest, is the paraphrase, and abbreviated (and often skillfully prepared) restatement of someone else's analysis or conclusion, without acknowledgment that another person's text has been the basis for the recapitulation."

The paragraphs above are from H. Martin and R. Ohmann, *The Logic and Rhetoric of Exposition, Revised Edition.* Copyright 1963, Holt, Rinehart and Winston.

## Academic Conduct Code

I. **Philosophy of Discipline**

The objective of Boston University in enforcing academic rules is to promote a community atmosphere in which learning can best take place. Such an atmosphere can be maintained only so long as every student believes that his or her academic competence is being judged fairly and that he or she will not be put at a disadvantage because of someone else's dishonesty. Penalties should

be carefully determined so as to be no more and no less than required to maintain the desired atmosphere. In defining violations of this code, the intent is to protect the integrity of the educational process.

## II. Academic Misconduct

Academic misconduct is conduct by which a student misrepresents his or her academic accomplishments, or impedes other students' opportunities of being judged fairly for their academic work. Knowingly allowing others to represent your work as their own is as serious an offense as submitting another's work as your own.

## III. Violations of this Code

Violations of this code comprise attempts to be dishonest or deceptive in the performance of academic work in or out of the classroom, alterations of academic records, alterations of official data on paper or electronic resumes, or unauthorized collaboration with another student or students. Violations include, but are not limited to:

A. **Cheating on examination**. Any attempt by a student to alter his or her performance on an examination in violation of that examination's stated or commonly understood ground rules.

B. **Plagiarism.** Representing the work of another as one's own. Plagiarism includes but is not limited to the following: copying the answers of another student on an examination, copying or restating the work or ideas of another person or persons in any oral or written work (printed or electronic) without citing the appropriate source, and collaborating with someone else in an academic endeavor without acknowledging his or her contribution. Plagiarism can consist of acts of commission-appropriating the words or ideas of another- or omission failing to acknowledge/document/credit the source or creator of words or ideas (see below for a detailed definition of plagiarism). It also includes colluding with someone else in an academic endeavor without acknowledging his or her contribution, using audio or video footage that comes from another source (including work done by another student) without permission and acknowledgement of that source.

C. **Misrepresentation or falsification of data** presented for surveys, experiments, reports, etc., which includes but is not limited to: citing authors that do not exist; citing interviews that never took place, or field work that was not completed.

D. **Theft of an examination**. Stealing or otherwise discovering and/or making known to others the contents of an examination that has not yet been administered.

E. **Unauthorized communication during examinations**. Any unauthorized communication may be considered prima facie evidence of cheating.

F. **Knowingly allowing another student to represent your work as his or her own**. This includes providing a copy of your paper or laboratory report to another student without the explicit permission of the instructor(s).

G. **Forgery, alteration, or knowing misuse of graded examinations, quizzes, grade lists, or official records of documents**, including but not limited to transcripts from any institution, letters of recommendation, degree certificates, examinations, quizzes, or other work after submission.

H. **Theft or destruction of examinations or papers** after submission.
I. **Submitting the same work in more than one course** without the consent of instructors.
J. **Altering or destroying another student's work or records**, altering records of any kind, removing materials from libraries or offices without consent, or in any way interfering with the work of others so as to impede their academic performance.
K. **Violation of the rules governing teamwork**. Unless the instructor of a course otherwise specifically provides instructions to the contrary, the following rules apply to teamwork: 1. No team member shall intentionally restrict or inhibit another team member's access to team meetings, team work-in-progress, or other team activities without the express authorization of the instructor. 2. All team members shall be held responsible for the content of all teamwork submitted for evaluation as if each team member had individually submitted the entire work product of their team as their own work.
L. **Failure to sit in a specifically assigned seat during examinations.**
M. **Conduct in a professional field assignment that violates the policies and regulations of the host school or agency.**
N. **Conduct in violation of public law occurring outside the University that directly affects the academic and professional status of the student, after civil authorities have imposed sanctions.**
O. **Attempting improperly to influence the award of any credit, grade, or honor.**
P. **Intentionally making false statements to the Academic Conduct Committee or intentionally presenting false information to the Committee.**
Q. **Failure to comply with the sanctions imposed under the authority of this code.**

## Disability Services

In accordance with University policy, every effort will be made to accommodate unique and special needs of students with respect to speech, hearing, vision, or other disabilities. Any student who feels he or she may need an accommodation for a documented disability should contact the Office of Disability Services at (617) 353-3658 or at access@bu.edu for review and approval of accommodation requests.

## Technical Support

**Experiencing issues with BU websites or Blackboard?**

It may be a system-wide problem. Check the BU Information Services & Technology (IS&T) news page for announcements.

Boston University technical support is available via email (ithelp@bu.edu), the support form, and phone (888-243-4596). Please note that the IT Help Center has multiple locations. All locations can be reached through the previously mentioned methods. For IT Help Center hours of operation please visit their contact page. For other times, you may still submit a support request via email, phone, or the support form, but your question won't receive a response until the following day. If you aren't calling, it is highly recommended that you submit your support request via the technical-support form as this provides the IS&T Help Center with the best information in order to resolve your issue as quickly as possible.

Examples of issues you might want to request support for include the following:

- Problems viewing or listening to sound or video files
- Problems accessing internal messages
- Problems viewing or posting comments
- Problems attaching or uploading files for assignments or discussions
- Problems accessing or submitting an assessment

To ensure the fastest possible response, please fill out the online form using the link below:

| **IT Help Center Support** |
| --- |
| 888-243-4596 or local 617-353-4357 or http://www.bu.edu/help/tech/learn |
| Check your open tickets using BU's ticketing system. |