

Course syllabus

I. Course Information

1. Course Name: R FOR BIOMEDICAL INFORMATICS

2. Course Number: RBIF 111

3. Course weeks: 10 course weeks, starting on Wednesdays: from 5/21/14 through 7/29/14. The course week runs from Wednesday through Tuesday. Week 10 concludes on 7/29/14.

4. Instructor's Name and Contact Information

Dr. Andrey Sivachenko

Email: asivache@brandeis.edu

Phone: 801-450-5934 (cell)

Office hours by appointment; preferred time: Mon, Tue 5-7pm; Wed through Fri. 7-9PM

5. Document Overview

This syllabus contains all relevant information about the course: its objectives and outcomes, the grading criteria, the texts and other materials of instruction, and of weekly topics, outcomes, assignments, and due dates.

Consider this your roadmap for the course. Please read through the syllabus carefully and feel free to share any questions that you may have. Please print a copy of this syllabus for reference.

6. Course Description

This course is an advanced mathematics and applied statistics course that will introduce students to data analysis methods and statistical testing. It provides a foundation for Biological Data Mining and Modeling (RBIF 112) and Design and Analysis of Microarray Experiments (RBIF114). This course will use R (a statistical programming language) to introduce students to descriptive and inferential statistics, basics of programming, common data structures and analysis techniques. It will cover methods important to data analysis such as t-tests, chi-squared analysis, Mann-Whitney tests, correlation and regression, ANOVA, PCA, tests of significance, Fisher's exact test, etc.

Prerequisites: IMPORTANT: it is assumed that the student has some introduction to probability and statistics at the prerequisite or undergraduate level. It is also assumed that students were previously exposed to programming/scripting. If you do not have an introduction to these courses, you must contact the instructor immediately on enrollment or before.

7. Materials of Instruction

a. Recommended Text (primary)

- J. Adler, R in a Nutshell
Publisher: O'Reilly Media
ISBN: 9780596801700

Note that the lectures are self-consistent and do not refer to the textbook directly. However, this is a very useful book to consult with.

b. Required Software and Tools

- R software, available from CRAN: <http://cran.r-project.org/>

c. Recommended Text(s) / Journals

- Michael Crawley: The R Book (hardcover). 950 pages
Publisher: Wiley (June 15, 2007)
ISBN-10: 0470510242
ISBN-13: 978-0470510247

This book is not required but may be helpful to you over the course of your career. It is a generalized 'cookbook' for R solutions.

- "An introduction to R" available in HTML (<http://cran.r-project.org/doc/manuals/R-intro.html>) or PDF (<http://cran.r-project.org/doc/manuals/R-intro.pdf>) formats.

d. Online Course Content

This is a Distance Learning (DL) course, which will be hosted at Brandeis' LATTE site, available at <http://latte.brandeis.edu>. The site contains the course syllabus, weekly topic notes, assignments, and discussion forums through which we will communicate during this course.

8. Overall Course Objectives

The course is intended to provide students with working knowledge of:

R Programming: basic and advanced analysis and graphing techniques
Approaches for data analysis with linear algebra
Statistical modeling
Formal inference and estimation
Contingency tables and Fisher's exact test
Confidence intervals and hypothesis tests
Correlation and regression
Multiple Linear Regression
Linear models
Analysis of variance
Non-parametric methods

9. Overall Course Outcomes

At the end of the course, students will be able to:

1. Write programs in R for data analysis

2. Discern appropriate data models to answer specific questions
3. Analyze data with a number of statistical models
4. Conduct explorative analysis of large data sets
5. Choose appropriate methods of analysis for a given problem
6. Perform tests to validate models against data
7. Evaluate various statistical methods on the basis of their strengths and weaknesses
8. Take advanced courses on the analysis of biological data (RBIF112 and RBIF114)

10. Course Grading Criteria

Percent	Component
40%	Homework 1-8, averaged
15%	Final examination
15%	Take-home midterm examination
30%	Class participation

1. Students will be assigned homework on a weekly basis: 8 homework assignments over the term.
2. Final exam: last week of the classes.
3. Mid-term examination: fifth week of the classes.
4. Class participation is graded based on student's replies to discussion questions and substantive replies to other posts on the class forum. Students are expected to read weekly course content and are required to
 - Post at least one original substantive response to each of the two discussion questions offered each week; the first response is due by Saturday midnight EST and the second response is due by Monday midnight EST
 - Post at least two substantive replies to the posts of others in the discussion question forums by Tuesday midnight EST
 - Make their required postings on at least three different days.

Participation Evaluation: Each week, 100 raw points may be earned toward the participation component of the grade.

Original Responses: Maximum raw points earned for original responses each week: 30 points each, 60 points total. Evaluation Criteria: Includes your own insights into the topics, sharing your professional experiences as appropriate and your own conclusions; Includes references to weekly required readings and/or other external sources, cited appropriately; Answers the question posed completely; Consists of at least 200-300 words; Well written, with no spelling or grammatical errors
Original responses one day late: -15 out of 30 possible raw points; more than one day late: no credit

Substantive Replies: Maximum raw points earned for substantive replies: 15 points each, 30 points total. Evaluation Criteria: Substantive (beyond an "I agree" or complimentary post) with follow-on points from your related experiences and/or from the readings and follow-up questions of others to extend the conversation (encouraged); consists of at least 200 words.
Replies one or more days late: no credit

Posting on Three Days of the Course Week

- Post on three days of the course week: 10 points (full credit)
- Post on two days of the course week: 6 points
- Post on one day of the course week: 2 points

II. Weekly Information

Week 1	5/21/2014
Outcomes	The students will be able to load, plot and save data in R, use common data structures, functions and operators, access R documentation.
Readings	Adler, Part 1 (Ch 1-4), Ch.12:pp147-161, Ch. 14: 211-243
Assignments	Homework 1

Week 2	5/28/2014
Outcomes	The students will be able to visualize data and inspect data plots to reveal patterns and irregularities in data; understand concept of null hypothesis, plan data analysis, use simple resampling simulation and t-test
Readings	Adler, Ch. 17; Ch.18: 341-347;
Assignments	Homework 2

Week 3	6/4/2014
Outcomes	The students will be able to fit simple linear models in R and test whether model assumptions are defensible; use probability distributions and random numbers.
Readings	Adler, Ch. 18: pp. 348-352; Ch. 20: pp.371-385
Assignments	Homework 3

Week 4	6/11/2014
Outcomes	Students will be able to use contingency tables and analysis of variance, calculate confidence intervals and interpret results of hypotheses' tests.
Readings	
Assignments	Homework 4

Week 5	6/18/2014
Outcomes	Students will be able to use permutation to test for significance, estimate confidence intervals via bootstrap and use priors for Bayesian analysis.
Readings	
Assignments	Take-home midterm examination, material covered: Week 1 through Week 4

Week 6	6/25/2014
Outcomes	Students will be able to fit single predictor regression model, interpret its diagnostics and use cross-validation to assess predictive accuracy of the model.
Readings	Adler 390-394
Assignments	Homework 5

Week 7	7/02/2014
Outcomes	The students will be able to fit linear models with multiple predictors, identify common problems with model fitting, assess and compare multiple models.
Readings	Adler 374-379
Assignments	Homework 6
Week 8	7/09/2014
Outcomes	The students will be able to use basic linear models with categorical variables and non-normal noise terms and do basic analysis of time series data.
Readings	Adler 394-399
Assignments	Homework 7
Week 9	7/16/2014
Outcomes	Advanced topics: the students will be able to analyze simple datasets using non-parametric methods, random effects and explore multivariate datasets.
Readings	Adler 355-358
Assignments	Homework 8 – preparation for final exam
Week 10	7/23/2014
Outcomes	Course wrap-up. Overview of good practices in data analysis. FINAL EXAM
Readings	Dupuy A, Simon RM. Critical Review of Published Microarray Studies for Cancer Outcome and Guidelines on Statistical Analysis and Reporting. JNCI (2007) 99: 147
Assignments	All material from the entire course will be covered in final exam.

III. Course Policies and Procedures

1. Late Policies

All homework (mid-term exam is homework!) is due at midnight on the last day of the course week (Tuesday).

Homework submitted after that time will be deducted 10% points for each day late.

2. Grading Standards

- All work will be expected to offer both computer programs executed as well as results, as required in the homework assignment document.
- No outside R programs may be used. Plagiarism is not allowed and University policies on plagiarism are strictly in force. If you have any questions on the appropriateness of a reference or a computer program you would like to use, please contact the instructors before handing your homework in or proceeding using others' work. It is better to be safe than sorry.

How Points and Percentages Equate to Grades:

100-94	A
93-90	A-
89-87	B+
86-83	B
82-80	B-
79-77	C+

76-73	C
72-70	C-
69-67	D+
66-63	D
62-60	D-
59 or <	F

Feedback

Feedback will be provided on assignments and the exam within 10 days of the due date. Within 7 days of the close of each week, feedback will be provided on weekly participation via the associated Weekly Participation Feedback assignments; you do not have to submit anything to these assignments. I will be recording your weekly participation grades and providing narratives describing your discussion posts (participation) for that week.

3. Workload Expectations

- Students are responsible to explore each week's materials and submit required work by their due dates. On average, a student can expect to spend approximately 2-4 hours per week reading and approximately 6-9 hours per week completing assignments and posting to discussions. The calendar of assignments and due dates is located at the end of this syllabus, and all assignments are due by the close of the associated week (Tuesday evenings).

4. Class Schedule

Class Schedule:

Week	Dates
1	May 22-28
2	May 29-June 4
3	June 5-11
4	June 12-18
5	June 19-25
6	June 26-July 2
7	July 3-9
8	July 10-16
9	July 17-23
10	July 24-30

5. Access to the Course Site

Members of the University's technical staff have access to all course sites to aid in course setup and technical troubleshooting. Program Chairs and a small number of Graduate Professional Studies (GPS) staff have access to all GPS courses for oversight purposes. Students enrolled in GPS courses can expect that individuals other than their fellow classmates and the course instructor(s) may visit their course for various purposes. Their intentions are to aid in technical troubleshooting and to ensure that quality course delivery standards are met. Strict confidentiality of student information is maintained.

IV. University and Division of Graduate Professional Studies Standards

Please review the policies and procedures of Graduate Professional Studies, found at <http://www.brandeis.edu/gps/students/studentresources/policiesprocedures/index.html>. Among them, we would like to highlight the following.

Learning Disabilities

If you are a student with a documented disability on record at Brandeis University and wish to have a reasonable accommodation made for you in this course, please contact me immediately.

Academic Honesty and Student Integrity

Academic honesty and student integrity are of fundamental importance at Brandeis University and we want students to understand this clearly at the start of the term. As stated in the Brandeis *Rights and Responsibilities* handbook, "Every member of the University Community is expected to maintain the highest standards of academic honesty. A student shall not receive credit for work that is not the product of the student's own effort. A student's name on any written exercise constitutes a statement that the work is the result of the student's own thought and study, stated in the student's own words, and produced without the assistance of others, except in quotes, footnotes or references with appropriate acknowledgement of the source." In particular, students must be aware that material (including ideas, phrases, sentences, etc.) taken from the Internet and other sources **MUST** be appropriately cited if quoted, and footnoted in any written work turned in for this, or any, Brandeis class. Also, students will not be allowed to collaborate on work except by the specific permission of the instructor. Failure to cite resources properly may result in a referral being made to the Office of Student Development and Judicial Education. The outcome of this action may involve academic and disciplinary sanctions, which could include (but are not limited to) such penalties as receiving no credit for the assignment in question, receiving no credit for the related course, or suspension or dismissal from the University.

Further information regarding academic integrity may be found in the following publications: "In Pursuit of Excellence - A Guide to Academic Integrity for the Brandeis Community", "*(Students') Rights and Responsibilities Handbook*" AND "Continuing Studies Student Handbook". You should read these publications, which all can be accessed from the Continuing Studies Web site. A student that is in doubt about standards of academic honesty (regarding plagiarism, multiple submissions of written work, unacknowledged or unauthorized collaborative effort, false citation or false data) should consult either the course instructor or other staff of the Rabb School for Continuing Studies.

University Caveat

The above schedule, content, and procedures in this course are subject to change in the event of extenuating circumstances.