

GNET743: Introduction to R and Statistics For Biomedical Scientists Spring 2014 Syllabus

Updated: 2014-03-19

Basic information

Course Identifiers: This document describes the syllabus for GNET743 in the Genetics Curriculum of the BBSP.

When: 10 am – 11:15 am, M/Tue/W.

Where: Lectures and computer lab in the Biogen Idec Classroom 307, Health Sciences Library.

Materials: All learning materials will be posted on “Sakai” at Sakai <https://sakai.unc.edu>.

Restrictions: Class is limited to twenty four (24) students.

Instructors

Lead instructor: Prof William Valdar, Room 5113, 120 Mason Farm Road, Genetic Medicine Building, Campus Box # 7264, Chapel Hill NC 27599. Tel: +1 919 843 2833. Email: william.valdar@unc.edu. Web: <http://valdarlab.unc.edu>

Co-instructor: Prof Leslie Lange, Room 5112, 120 Mason Farm Road, Genetic Medicine Building, Campus Box # 7264, Chapel Hill NC 27599. Tel: +1 919 966 9562. Email: leslie_lange@med.unc.edu. Web: <http://genetics.unc.edu/faculty/leslie-lange>

Teaching Assistant: Ms Bailey Peck <bpeck@email.unc.edu>. Office hours for the TA will be arranged and posted at the beginning of the first class and updated as necessary.

Target Audience

This course is targeted at graduate students in the biomedical sciences who are working in experimental laboratories but must perform quantitative analyses or visualize quantitative data at some point in their research. It has moderate emphasis on examples found in genetics and genetic epidemiology, but is otherwise general, using subject non-specific examples to provide intuition about statistical techniques commonly used in biomedical research. In addition to biomedical science students, it may also be of interest to graduate students in related disciplines, including epidemiology and health sciences.

Course Prerequisites

The course is open to all graduate students of the Biological and Biomedical Sciences Program (BBSP) at UNC Chapel Hill. Other students, staff, or faculty may attend for credit, on an auditor basis or informally **only if**

- They have prior permission from the lead instructor, and
- There is space: that is, if they are not taking up a spot that would be otherwise used by a non-auditing (ie, full credit) BBSP student.

Course Goals and Key Learning Objectives

This course will introduce the free and increasingly popular statistical analysis and graphics package R, and teach fundamental statistical concepts students are likely to encounter in biomedical research. The course has no formal requirements, but it is hoped that by the end students will have made progress in the following areas:

- 1) Using R as a calculator
- 2) Making simple plots
- 3) Building up simple plots into sophisticated plots and graphics
- 4) Manipulating, filtering and merging (potentially) large datasets
- 5) Familiarity with the linear model, in particular
 - a. understanding the meaning of parameters
 - b. experience in fitting batch effects
 - c. understanding the concepts of confidence intervals (CIs) and how this relates to variability and uncertainty
 - d. understanding the basics of hypothesis testing
- 6) Exposure to concepts underpinning logistic regression, t-tests, ANOVA.
- 7) Exposure to Principal Components Analysis (PCA)

Core competencies: making graphs in R, data manipulation, basic statistical analysis, interpretation of results from linear regression analyses.

Course Requirements

To obtain full credit, students must attend at least 80% of the lectures, complete all four homeworks, and achieve at least a passing overall grade.

Dates

Homework assignments will typically be distributed on Wednesdays after class, with a deadline for electronic submission at least a week later, typically 5pm on the Thursday of the following week. Anonymous student evaluations, required for 5% of the course marks, will be distributed for completion on Sakai within approximately a week of course completion. Students will have a week to complete the student evaluation.

Grades

Grades for the course (F,L,P,H) will be based on performance in the homeworks and on completion of the course evaluation. Specifically, the homeworks collectively account for 95% of the course marks, and completion of the anonymous evaluation accounts for the remaining 5%. Each homework will include multiple questions each providing a stated maximum number of points. The total number of points achieved by a student divided by the total possible will be scaled to the range 0 to 95 and used as the percentage of the grade arising from coursework. There is **no final exam**.

Course Policies

Students must attend the entire duration of at least 80% of the lectures unless they have permission of the lead instructor to do otherwise. Students are expected to be prompt, polite, collaborative when (and only when) asked, and to answer questions in class. Failure to hand in a homework on time without reasonable justification (eg, sickness) will result in automatic loss of 10% of that homework's maximum allowable points for each day over the deadline.

Course Resources:

No textbooks are required, but students are encouraged to seek out appropriate tutorials and other reference material on the internet. A recommended textbook considered by the instructors to contain useful supporting material is Verzani J (2004) "Using R for Introductory Statistics" Chapman and Hall/CRC press.

Honor Code:

Students may collaborate in class, but each student's homework should be their own. In completing the homework, however, students are nonetheless encouraged to consult the lecture notes, online material, books and any other "passive" sources. They may discuss general strategies and concepts in R with their classmates and with the TA, and may ask the TA for clarification about the content of questions. The TA may provide guidance as to where they might be able to find example material that addresses problems similar (but not identical) to those posed in the homework.

Time Table

Lecture	Date	Day	Instructor	Topic
Week 1: R				
1	17-Feb	M	Valdar	Using R: a one session tutorial
2	18-Feb	Tu	Valdar	Functions and objects 1: functions, plotting, primitive data types
3	19-Feb	W	Valdar	Functions and objects 2: primitive data types, 2d data structures, factors
HW	19-Feb	W	Valdar	Homework 1
Week 2: Linear models concepts				
4	24-Feb	M	Valdar	Linear models (pt 1)
5	25-Feb	Tu	Valdar	Linear models (pt 2)
6	26-Feb	W	Valdar	Significance testing
HW	26-Feb	W	Valdar/Lange	Homework 2
Week 3: Linear models in practice				
8	3-Mar	M	Lange	Predictors: confounding, interactions (pt 1)
9	4-Mar	Tu	Lange	Predictors: confounding, interactions (pt 2)
10	5-Mar	W	Lange	T-tests, correlations, non-parametric tests (pt 1)
HW	5-Mar	W	Lange/Valdar	Homework 3
Spring Break				
Week 4: Special topics				
10	17-Mar	M	Lange	T-tests, correlations, non-parametric tests (pt 2)
11	18-Mar	Tu	Lange	Undertaught fundamentals: causal inference, outliers, transformations, missingness
12	19-Mar	W	Valdar	Special topics: for loops, logistic regression, PCA
HW	19-Mar	W	Valdar	Homework 4

Syllabus Changes

The lead and/or co-instructors reserve to right to make changes to the syllabus, including project due dates.