

Special Topics in Applied Biostatistical Practice

Public Health 298 [CCN 34626], Spring 2017
Suzanne Dufault, Lucia Petito, Courtney Schiffman, Kelly Street
Instructor of Record: Sandrine Dudoit

Facilitator contact info

Lucia Petito lucia.petito@berkeley.edu	Kelly Street kstreet@berkeley.edu
Courtney Schiffman courtneys@berkeley.edu	Suzanne Dufault sdufault@berkeley.edu

Class info

Meeting place: 150D Moffitt, Mondays 5-6pm
Office hours: 114 Haviland Hall, Fridays 10-11am
<https://piazza.com/berkeley/spring2017/ph298/>

Prerequisites

Some statistical theory and computing, like
Stat 134/135/201A/201B/133/243.

Course description

This course is a rigorous presentation of the practical issues encountered when applying statistical methods to public health and biological data. Course material includes software to enable collaborative research, perform advanced statistical inference, and implement computational algorithms. All topics covered incorporate hands-on data analysis with computer software (primarily R).

Class will meet for one hour every week. A long-term project will be assigned mid-semester to give students an opportunity to apply appropriate methods to a data analysis of their choosing. Credit is available exclusively on an S/U basis.

Course learning objectives

By the end of this course, students will be able to:

- Facilitate collaborative research with tools such as knitr and GitHub.
 - Implement parallelized code for efficient computation
 - Visually communicate statistical findings, through use of effective graphics in R (using the packages lattice, ggplot2, maps)
 - Implement multiple imputation procedures in R (write code themselves, use packages mi and mice)
 - Summarize large data sets with dimensionality reduction techniques
 - Build and assess effective prediction models through bagging and boosting
-

Reference texts

All required readings will be provided to students as PDFs through the course website.

Some supplemental texts that might be of interest are:

- Rubin, D.B. (1987) *Multiple Imputation for Nonresponse in Surveys*. Wiley, New York.
- Lumley, T. (2010) *Complex Surveys: A Guide to Analysis Using R*. Wiley, New York
- Wickham, H. (2009) *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York.
- Tibshirani text

R Software

All analyses in this class will be done in R, a set of open-source statistical software. R is free, and can be downloaded at <http://www.r-project.org>. A basic familiarity with R is expected of all students in the class.

Evaluation procedures

To earn a pass, a student must:

- Attend all class meetings (exceptions will be considered on a case-by-case basis)
- Participate in every class discussions and workshops
- Design and execute a data analysis pipeline and prepare a brief (5-10 minute) presentation. Students may work alone or in pairs.

Additional policies

Personal computer use is encouraged, especially during the in-class workshops. Cell phone use is discouraged, but not banned. There will be no extra credit offered.

Academic integrity is a requirement for this class. For more information on the official university policy, see http://sa.berkeley.edu/sites/default/files/UCB-Code-of-Conduct-new%20Jan2012_0.pdf

Students with disabilities should first register with the Disabled Students' Program and then arrange a private meeting with the course facilitators to make particular arrangements. See <http://www.dsp.berkeley.edu> for more information.

Class schedule

Week	Date	Topic	Instructor
1	1/23	Intro/using the cluster	Kelly
2	1/30	Parallelization	Courtney
3	2/6	knitr/LaTeX/GitHub/RProjects	Suzanne
4	2/13	Linear Models / ANOVA	Courtney
5	2/20	*** ACADEMIC HOLIDAY ***	
6	2/27	PCA and Dimensionality Reduction	Kelly
7	3/6	Variable Importance and Selection	Courtney
8	3/13	Multiple Imputation	Lucia
9	3/20	Inverse Probability Weighting	Lucia
10	3/27	*** SPRING BREAK ***	
11	4/3	Prediction: Bagging and Boosting	Kelly
12	4/10	Categorical Data Analysis	Suzanne
13	4/17	Data Visualization	Kelly
14	4/24	GEE and Mixed Effects Modeling	Suzanne
15	5/1	RRR Week, digit recognition presentations	