The University of California - Los Angeles
Department of Statistics

Social Statistics
Statistics 216
Professor: Mark S. Handcock

Syllabus and Description

Professor:
Mark S. Handcock

Office Hours:
see the Bruin Learn *Home* page
held in-person and on zoom simultaneously
Other times by arrangement. Clearly composed email questions sent to the will receive written replies

Motivation and Synopsis

Statistics C116/C216 is a second course in social statistics and will *focus on Bayesian statistical analysis*. It will cover the principles of Bayesian statistics, Bayesian data analysis and modeling. In particular it will develop the ideas in the context of linear regression, its non-linear generalizations, and hierarchical models. While the idea and principles are general, all applications and models will be chosen from those most relevant to the social sciences.

This course is designed for graduate students majoring in Statistics, the Social Sciences and advanced undergraduate students who are planning to attend graduate school.

This course is most appropriate for student seeking additional training in the application of Bayesian statistics to data and are looking for an introductory course that advances rapidly. The ultimate goal is to equip students with the analytical and programming skills necessary to address social statistics problems within the Bayesian paradigm based on current data and technologies.

The course has various perspectives:

- It focuses on conceptual understanding of the Bayesian statistics.
- It focuses on conceptual understanding of the the primary models used in social statistics.

- It involves the analysis of real-data.
- It involves implementing the methods using freely available software.

The course will involve the practical application of the ideas of statistical computing and their implementation through statistical software, particularly `R`. As statistical computation is essential for many of the modeling approaches, expertise will need to be developed.

Prerequisites

- At least two courses in statistics, one of which is on regression modeling.
- People may still take this class as long as they have background on matrix algebra, probability theory, and programming skills. To do this attend the first classes and we can assess if this is advisable.

Structure of the Course

There will be two lectures per week.

Textbooks

[**H**] Hoff, Peter D.
A First Course in Bayesian Statistical Methods (2009).
Springer
Required and available online as a ebook from the library, free for UCLA students.
`https://link.springer.com/book/10.1007%2F978-0-387-92407-6`

[**GH**] Gelman, Andrew and Hill, Jennifer
Data Analysis Using Regression and Multilevel/Hierarchical Models (2007).
Cambridge University Press
Recommended
`http://www.amazon.com/Analysis-Regression-Multilevel-Hierarchical-Models/dp/052168689X`

Students must be connected to the UCLA network to obtain their free download. Students who would like to download the textbook off-campus may do so by connecting the the UCLA network via VPN `https://www.bol.ucla.edu/services/vpn/all.html`.

In addition to these books, there are a multitude of books covering pieces of the course content and from varying perspectives. I suggest you use one or more of the following.

Background on `R`

[**W**] Wickham, Hadley.
Advanced R. (2014).
Chapman and Hall/CRC.
Required and available online as a ebook from the library, free for UCLA students.
`http://www.crcnetbase.com/doi/book/10.1201/b17487`

[**M**] Maillardet, Robert, Owen Jones, and Andrew Robinson.
Introduction to Scientific Programming and Simulation Using R. (2014).
Chapman and Hall/CRC.
Required and available online as a ebook from the library, free for UCLA students.
`http://www.crcnetbase.com/doi/book/10.1201/9781420068740`

Other Resources

You can read the other books with different perspectives online for free from any UCLA account, starting from:
`proquest.safaribooksonline.com/search?q=BOOKTITLE%20r`

Syllabus of the Course

The syllabus of the course will develop on the following weekly schedule. The some later topics may not be reached and we will make choice among them toward the end of the quarter.

Chapter	Contents discussed
1	Introduction and examples
	– Why the Bayesian paradigm?
	– Where we are going
2	Belief, probability and exchangeability
	– Notions of Subjective Probability
3	One-parameter models
4	Statistical Decision Theory
5	Introduction to Monte Carlo Methods
	– Simulation, estimation, sampling, optimization
6	An important case: The Gaussian model
7	Posterior approximation with the Gibbs sampler
8	The multivariate normal model
9	Group comparisons and hierarchical modeling
10	Linear regression
11	Generalized linear models
	– Nonconjugate priors and Metropolis-Hastings algorithms
12	Multilevel Modeling in Stan and R
13	Linear and generalized linear mixed effects models
14	Latent variable methods for ordinal data

Learning Outcomes

- Students will explain the concept of subjective probability.
- Students will be able to formulate simple stochastic models for social phenomena.
- Students will be able to explain Bayesian inferential procedures for stochastic models for social phenomena.
- Students will explain the difference between Bayesian and Frequentist statistical analyses.
- Students will be able to write code in Stan and R to implement Bayesian inference.

Course Webpage and Discussion Forum

The course has a webpage through the UCLA *Bruin Learn*, `https://bruinlearn.ucla.edu/`. The webpage will be continuously updated throughout the course with handouts, homework assignments and solutions. Users sign in to Bruin Learn with their UCLA Logon IDs.

I will be using Bruin Learn Discussion Board to provide discussion of issues in class and related questions. For questions that might be of interest to other students, please use Bruin Learn rather than solely emailing me. There other students and the TA can answer questions in addition to me. Example of questions are about interesting articles you have seen in the media, problems with access to resources, homework or computer questions. Enjoy!

Please regularly consult this class's Bruin Learn Bruin Learn home page, and the archive of the Announcements mailing list. It will contain lecture notes, homework, solutions and course information.

Computer Usage and Software

The computer is the scientific laboratory of the applied researcher in quantitative fields. As such this course requires the student to develop a degree of comfort and competence "in the lab".

For those newish to `R`, a good computer interface to `R` is the RStudio IDE, which you can download from `www.rstudio.com`.

Course Requirements and Grades

- 50% Homeworks (5 or 6, weighted equally, none are dropped)
- 50% Final project (written report)

Homework

There will be regular homeworks on the theory, modeling and data analysis Students will be graded on a scale of 1 to 10 for each homework.

None of the scores will be dropped. It is your responsibility to verify that your work successfully uploaded by the deadline. All homework assignments will be posted. Students will submit the solution files electronically via *gradescope* from the Vruin Learn *Assignments*

page. A PDF file is preferred and an `R` code file can additionally be uploaded. The use of `R` Markdown and/or `knitr` is suggested.

Students are free to discuss homework problems and solutions. Discussing the contents of the course with fellow students can be a valuable element of the learning process, and doing so is therefore generally encouraged. However, each student must hand in their own solutions, and the student should, if asked, be able to explain the solutions.

The project will be to undertake an Bayesian analysis of data that you find interesting. You can select any data set you like, but preferable related to your graduate work or thesis area. I do not want a quick and routine analysis; a good job will show understanding of the problem and possible solutions and techniques to consider. The technical results should be stated clearly. The report must contain a clearly written conclusion section giving a non-technical summary that is concise and informative.

Unless approved by me, the data set should contain at least 60 observations, at least one categorical variable, at least one continuous variable, and some hierarchical structure. One of the variables should be such that it would be of interest to try to model it or predict it (e.g., party voted for, son's occupation). Do *not* merely use data from a textbook the world is an interesting place! All data sources must be cited, and described.

For details of the project, see the "Homework" section.

Late Policy for Homework (silly but necessary)

There is a 10 minute grace period. Submissions up to 10 minutes late will be accepted with no penalty.

Homework assignments submitted 11 minutes late or more will be accepted with penalty. There is a minimum deduction of 1 point for being less than one hour late. An additional 0.5-point deduction will be taken for each additional hour it is late.

Thus, an assignment that is between 11 and 59 minutes late will receive a 1-point deduction. An assignment that is between 1 hour and 1:59 late will receive a 1.5-point deduction. An assignment that is between 2 hours and 2:59 late will receive a 2-point deduction, and so on.

I strongly advise uploading the homework to the CCLE well in advance of the deadline in case there are connectivity problems or server issues.

Academic Integrity

As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. All students must uphold University of California Standards of Student Conduct as administered by the Office of the Dean of Students. Students are subject to disciplinary action for several types of misconduct, including but not limited to: cheating, multiple submissions, plagiarism, prohibited collaboration, facilitating academic dishonesty, or knowingly furnishing false information. You may have assignments or projects in which you work with a partner or with a group. For example, you are welcome, and even encouraged, to work with others to solve homework problems. Even though you are working together, the assignment

you submit for a grade must be IN YOUR OWN WORDS, unless you receive specific instructions to the contrary. For more information about academic integrity, please go to www.deanofstudents.ucla.edu.

I welcome comments or suggestions about the course at any time, either in person, by letter, or by email. Please feel free to use these ways make comments to me about any aspect of the course.

Support

Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. Students who have experienced sexual harassment or sexual violence can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 1st Floor Wooden Center West, CAREadvocate@caps.ucla.edu, (310) 206-2465. You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, Kathleen Salvaty, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417.

I welcome comments or suggestions about the course at any time, either in person, by letter, or by email. Please feel free to use these ways make comments to me about any aspect of the course.