BIOSTAT698/EPID815: Modern Statistical Methods in Epidemiologic Studies

Tue, Thu/8:30 – 10 AM/M4318 SPH-2

Professor: Veronica Berrocal
Biostatistics
SPH-II M4525, 734-763-5965
berrocal@umich.edu
Office Hours: WED 10AM – 11AM

Sung Kyun Park
Epidemiology
SPH-II M5541, 734-936-1719
sungkyun@umich.edu
Office Hours: WED 11AM – 12PM

Course Description: The goal of this pilot course is to create an interdisciplinary educational experience for Ph.D. students in Epidemiology (also available as an optional elective for Masters students in Biostatistics) through a uniquely designed course that contains lectures on advanced biostatistical methods, but places them in the context of epidemiological applications.

Course Materials: This is an advanced methods course. The instructors will provide comprehensive lecture notes via Canvas. Papers identified for required and recommended reading will be available in Canvas. Reading assignments will involve both biostatistical methodology papers/books as well as readings from the epidemiologic literature related to the outcomes and risk factors that we will consider in each project. A tentative list is provided at the end of this syllabus.

Pre-requisites: Epidemiology doctoral students: EPID 600 (or equivalent), BIOSTAT 522 and BIOSTAT 523 (or equivalent), Experience using R required (at least the levels covered in EPID 798 (Epidemiologic Data Analysis using R or equivalent) required. BIOSTAT 512 (Longitudinal Analysis) recommended.
Biostat Students: EPID 503 (or equivalent), BIOSTAT 650 and 651 required. Experience using R required. Biostat 695 recommended.

Course Goals: Students enrolled in the class will learn about cutting edge statistical methods in these three contemporary topics that arise frequently in the present scientific context. These three topics are: (a) Variable selection and Statistical Learning; (b) Bayesian statistics; and (c) Spatial modeling for areal data. The course will equip the new generation epidemiologists with state-of-the-art statistical methods in these domains, and teach them the craft of translating a practical problem into mathematical equations. However, the entire theoretical learning process will be placed in the context of sophisticated modeling of data from large complex studies with a focused problem to solve.
Competencies: After completing this class, students are expected to be able to attain the following competencies:

Biostatistics Competencies:
- Develop knowledge to communicate and collaborate effectively with scientists in a variety of health-related disciplines to which biostatistics are applied (e.g. public health, medicine, genetics, biology; psychology; economics; management and policy).
- Become well-versed in the application of core statistical techniques (biostatistical inference, linear regression, generalized linear models, nonparametric statistical methods, linear mixed models) and 4-5 selected statistical specialization techniques.
- Select appropriate techniques and apply them to the processing of data from health studies.
- Interpret the results of statistical analysis and convert them into a language understandable to the broad statistical community.
- Develop written and oral presentation skills and other scientific reporting skills, based on statistical analyses for public health, medical and basic scientists and educated lay audiences.

Epidemiology Competencies
- Employ state-of-the-art statistical and other quantitative methods in the analysis of epidemiologic data.
- Demonstrate a thorough understanding of causal inference, sources of bias, and methods to improve the validity of epidemiologic studies.
- Provide clear and effective oral communications of epidemiologic concepts, methods, results, and implications to scientists, students, policy makers, and the public.

Course Requirements: The course will consist of three projects related to the three modules. All three projects will be in group. Groups will be assigned in each project by instructors. The grading scheme will be the following:

Course Requirements: The course will consist of three projects related to the three modules. All projects will be in group and group members will change project to project. The grading scheme will be the following:

- 3 written reports: 25% × 3 = 75%
- 3 group presentations: 5% × 3 = 15%
- Class participation: 10%

The written report should contain an introduction that gives the background of the problem, a detailed description of the model and methods used, analysis of results and a conclusion summarizing what was learned from this study, with a limit of 12 pages (single spaced) including tables and figures. The final report will be graded based on (1) clarity of presentation and organization, (2) accuracy of used methods, (3) thoughtfulness and creativity.

Classroom Expectations/Etiquette:
Class attendance will be required. Laptops are allowed for class purposes but no smart phones will be allowed in class. University policy specifies that students are responsible for all official correspondences sent to their standard University e-mail address. Students should check this.
account frequently. Group work and didactic discussions are a key element of this class. Students will be expected to engage and participate fully.

**Academic Integrity:**
The faculty and staff of the School of Public Health believe that the conduct of a student registered or taking courses in the School should be consistent with that of a professional person. Courtesy, honesty, and respect should be shown by students toward faculty members, guest lecturers, administrative support staff, community partners, and fellow students. Similarly, students should expect faculty to treat them fairly, showing respect for their ideas and opinions and striving to help them achieve maximum benefits from their experience in the School.

Student academic misconduct refers to behavior that may include plagiarism, cheating, fabrication, falsification of records or official documents, intentional misuse of equipment or materials (including library materials), and aiding and abetting the perpetration of such acts.

Please visit [http://www.sph.umich.edu/academics/policies/conduct.html](http://www.sph.umich.edu/academics/policies/conduct.html) for the full SPH Code of Academic Integrity and further definition of these terms.

**Student Well-being:**
SPH faculty and staff believe it is important to support the physical and emotional well-being of our students. If you have a physical or mental health issue that is affecting your performance or participation in any course, and/or if you need help connecting with University services, please contact the instructor or the Office of Academic Affairs.

Please visit [http://www.sph.umich.edu/students/current/#wellness](http://www.sph.umich.edu/students/current/#wellness) for more information.

**Student Accommodations:**
Students should speak with their instructors before or during the first week of classes regarding any special needs. Students can also visit the Office of Academic Affairs for assistance in coordinating communications around accommodations.

Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit [http://ssd.umich.edu/accommodations](http://ssd.umich.edu/accommodations) for more information on student accommodations.

Students who expect to miss classes, examinations, or other assignments as a consequence of their religious observance shall be provided with a reasonable alternative opportunity to complete such academic responsibilities. It is the obligation of students to provide faculty with reasonable notice of the dates of religious holidays on which they will be absent. Please visit [http://www.provost.umich.edu/calendar/religious_holidays.html#conflicts](http://www.provost.umich.edu/calendar/religious_holidays.html#conflicts) for the complete University policy.
Course Topics/Reading List:

The following reading list and schedule is targeted towards the three projects that will be assigned. For Fall 2016 we have identified the outcome and risk factors we will study in each module. We provide a brief summary of the data we are going to use to motivate the context for the assigned readings. This list may be updated later.

Reading List

Module 1: Variable Selection and Statistical Learning

Project-Related Paper


Required Reading

- Taylor and Tibshirani. Statistical Learning and Selective Inference. PNAS 2015;112:7629-7634.

Recommended Readings:

Regularized Regression (ridge regression, LASSO, elastic net)


Dimension Reduction (principal component analysis)


Machine Learning (regression-tree)


Bayesian Model Averaging


Applications


Module 2: Bayesian Statistics

Project-Related Paper

Required Reading

Recommended Readings:

Applications

Module 3: Spatial Modeling for Areal Data

Project-Related Paper

Required Reading

Recommended Readings:

Applications
• Xia, H., Carlin, B. P., and Waller, L. Hierarchical models for mapping Ohio lung cancer rates. Environmetrics 1997;8:107-120.
## Course Schedule

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
<th>Lecture</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Module 1</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>09/06/16</td>
<td>Course overview, Introduction to Project 1</td>
<td>SKP/VB</td>
</tr>
<tr>
<td>09/08/16</td>
<td>Model building (best subset, stepwise, forward, backward)</td>
<td>SKP</td>
</tr>
<tr>
<td>09/13/16</td>
<td>Ridge regression, LASSO, elastic net</td>
<td>BM</td>
</tr>
<tr>
<td>09/15/16</td>
<td>Non-parametric machine learning methods (CART, random forest)</td>
<td>BM</td>
</tr>
<tr>
<td>09/20/16</td>
<td><strong>Interim Presentation</strong></td>
<td>VB</td>
</tr>
<tr>
<td>09/22/16</td>
<td>Dimension reduction, principal component analysis</td>
<td>VB</td>
</tr>
<tr>
<td>09/27/16</td>
<td>Cross-validation, Bayesian model averaging</td>
<td>VB</td>
</tr>
<tr>
<td>09/29/16</td>
<td>Prediction models, integrated measures</td>
<td>SKP</td>
</tr>
<tr>
<td>10/04/16</td>
<td>Epidemiologic perspectives</td>
<td>SKP</td>
</tr>
<tr>
<td>10/06/16</td>
<td><strong>Final Presentation</strong></td>
<td>VB</td>
</tr>
<tr>
<td><strong>Module 2</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/11/16</td>
<td>Introduction to Project 2</td>
<td>SKP/VB</td>
</tr>
<tr>
<td>10/13/16</td>
<td>Overview of Bayesian Statistics</td>
<td>SKP</td>
</tr>
<tr>
<td>10/18/16</td>
<td><strong>Study Break (no class)</strong></td>
<td>VB</td>
</tr>
<tr>
<td>10/20/16</td>
<td>Introduction to Markov Chain Monte Carlo</td>
<td>VB</td>
</tr>
<tr>
<td>10/25/16</td>
<td>(MCMC) Markov Chain Monte Carlo: convergence assessments</td>
<td>VB</td>
</tr>
<tr>
<td>10/27/16</td>
<td>Bayesian linear model</td>
<td>VB</td>
</tr>
<tr>
<td>11/01/16</td>
<td><strong>Interim Presentation</strong></td>
<td>VB</td>
</tr>
<tr>
<td>11/03/16</td>
<td>Bayesian generalized linear model</td>
<td>VB</td>
</tr>
<tr>
<td>11/08/16</td>
<td><strong>Final Presentation</strong></td>
<td>VB</td>
</tr>
<tr>
<td><strong>Module 3</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11/10/16</td>
<td>Introduction to Project 3</td>
<td>SKP/VB</td>
</tr>
<tr>
<td>11/15/16</td>
<td>Overview of Spatial Statistics</td>
<td>VB</td>
</tr>
<tr>
<td>11/17/16</td>
<td>Moran’s I, Correlogram, Smoothing</td>
<td>VB</td>
</tr>
<tr>
<td>11/22/16</td>
<td>Conditionally AutoRegressive (CAR) models</td>
<td>VB</td>
</tr>
<tr>
<td>11/24/16</td>
<td><strong>Interim Presentation</strong></td>
<td>VB</td>
</tr>
<tr>
<td>11/29/16</td>
<td><strong>Thanksgiving Break (no class)</strong></td>
<td>VB</td>
</tr>
<tr>
<td>12/01/16</td>
<td>Other smoothing models: proper CAR models, simultaneous autoregressive models (SAR)</td>
<td>VB</td>
</tr>
<tr>
<td>12/06/16</td>
<td><strong>Final Presentation</strong></td>
<td>VB</td>
</tr>
<tr>
<td>12/08/16</td>
<td>Epidemiologic perspectives, Wrap-up</td>
<td>SKP</td>
</tr>
</tbody>
</table>