SYLLABUS

Primary Instructor:
Sung Kyun Park, Sc.D., M.P.H.
Office: SPH-II M5541
Tel: 734-936-1719
Email: sungkyun@umich.edu
Office Hours: by appointment

Guest Lecturers:
Brisa Sanchez, Ph.D. (Biostatistics)
Carina Gronlund, Ph.D. (EHS)

Course Time & Location: Fridays 9am–12pm, SPH Tower 2615

Format: Lecture + computing lab

Grading: homework + class projects

Recommended Textbooks:

R manuals (available in CTools and R website (www.r-project.org)):
- Chongsuvivatwong V. Analysis of Epidemiological Data Using R and Epicalc.

Prerequisite: BIOSTAT 560 and EPID 503 or 600.

Course goals: The goal of this course is to provide students with a set of new and advanced data analysis tools, especially unique to environmental epidemiologic research, using R statistical software.

Course description: This course is an applied epidemiologic data analysis using R statistical software, a free software environment for statistical computing and graphics, and especially covers unique features of environmental exposure data. The course will introduce parametric and non-parametric smoothing methods, such as natural splines, penalized splines, and locally weighted polynomial regression (LOESS), and how to implement them using generalized linear models (glm) and generalized additive models (gam). The course also covers data analysis subject to the dependence issue, such as longitudinal data and complex sampling data (e.g., NHANES) and time-series analysis in air pollution health effects that are widely used in environmental epidemiology. This course will provide an opportunity to analyze actual population data to learn how to model environmental epidemiologic data, and is designed particularly for students who pursue environmental epidemiologic research. The course will consist of lectures and hands-on practices in computer labs, homework assignments and final projects.

Acknowledgment: I would like to thank Drs. Joel Schwartz and Antonella Zanobetti of the Harvard School of Public Health for sharing their course notes and data.
Learning Objectives

<table>
<thead>
<tr>
<th>Learning objective #</th>
<th>Learning objective</th>
<th>Expected level of knowledge</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The following learning objectives are achieved through (check boxes):

- Lecture presentations
- Prescribed readings
- Tutorials with instructor or GSI
- Research assignments with term papers
- Research assignments with presentations
- Other (specify) Hands-on exercises in computer labs

The students taking this class are expected to learn about:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>Selection of statistical methods by variable characteristics</td>
<td>Intermediate</td>
</tr>
<tr>
<td>L2</td>
<td>The basics of R statistical program</td>
<td>Intermediate</td>
</tr>
<tr>
<td>L3</td>
<td>Parametric and nonparametric smoothing methods and generalized additive model (GAM)</td>
<td>Advanced</td>
</tr>
<tr>
<td>L4</td>
<td>Analysis of multi-level data using mixed effects models</td>
<td>Advanced</td>
</tr>
<tr>
<td>L5</td>
<td>Analysis of complex sampling survey data</td>
<td>Advanced</td>
</tr>
<tr>
<td>L6</td>
<td>Application of GAM in air pollution time-series study</td>
<td>Advanced</td>
</tr>
</tbody>
</table>
## Course Competencies

<table>
<thead>
<tr>
<th>Comp #</th>
<th>Competency</th>
<th>Expected level of competency</th>
</tr>
</thead>
<tbody>
<tr>
<td>C1</td>
<td>Select an appropriate statistical method depending on variable characteristics</td>
<td>Intermediate</td>
</tr>
<tr>
<td>C2</td>
<td>Perform statistical analyses using R</td>
<td>Intermediate</td>
</tr>
<tr>
<td>C3</td>
<td>Analyze non-linearly associated variables without categorizing them</td>
<td>Advanced</td>
</tr>
<tr>
<td>C4</td>
<td>Understand multi-level data and analyze using mixed effects models</td>
<td>Advanced</td>
</tr>
<tr>
<td>C5</td>
<td>Understand complex sampling design and analyze survey data using survey procedures</td>
<td>Advanced</td>
</tr>
<tr>
<td>C6</td>
<td>Understand what the time-series analysis widely used in air pollution epidemiology is and have a potential to perform it</td>
<td>Advanced</td>
</tr>
</tbody>
</table>

## Performance Evaluations

<table>
<thead>
<tr>
<th>Evaluation Method</th>
<th>Style or Content</th>
<th>Purpose</th>
<th>% of grade</th>
<th>Learning Objectives Satisfied</th>
<th>Competencies Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>Homework Assignment Paper</td>
<td>Paper</td>
<td>Assessment of student performance in each topic</td>
<td>60% (4 HWs: each 15%)</td>
<td>L1 to L4</td>
<td>C1 to C4</td>
</tr>
<tr>
<td>Final Paper</td>
<td>Paper</td>
<td>Evaluation of overall performance</td>
<td>40%</td>
<td>L1 to L6</td>
<td>C1 to C6</td>
</tr>
<tr>
<td>Date</td>
<td>Topics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Week 1 (01/10/14) | Lecture: Course introduction, Classification of variables, choice of statistical analysis based on variable characteristics  
Lab: Introduction to R                                                                                                                                 |
| Week 2 (01/17/14) | Lab: Simple descriptive data functions in R  
Packages ‘epicalc’ and ‘ggplot2’                                                                                                                                                                         |
| Week 3 (01/24/14) | Lecture: Linear regression models (confounding and model development, residuals and regression diagnostic, plotting covariate-adjusted association)  
Lab: Continue Simple descriptive data functions in R and LM function in R                                                                                                                               |
| Week 4 (01/31/14) | Lecture: Effect modification  
Lab: Homework 1 review and Continue LM function in R (HOMEWORK 1 DUE)                                                                                                                                 |
| Week 5 (02/07/14)  | Lecture: Statistical issues in biomarker data (Standardized vs. Adjusted)  
Lab: Continue LM function in R                                                                                                                                                                        |
| Week 6 (02/14/14)  | Lecture: Generalized linear model  
Lab: GLM function in R                                                                                                                                                                                       |
| Week 7 (02/21/14)  | Lecture: Introduction to scatterplot smoothing (by Brisa Sanchez)  
Lab: Parametric and non-parametric smoothing in R, methods for smoothing (regression splines, penalized splines) using package ‘splines’ (HOMEWORK 2 DUE)                                                    |
| Week 8 (02/28/14)  | Lecture: Generalized additive model  
Lab: GAM function in R, packages ‘gam’ and ‘mgcv’                                                                                                                                                       |
| Week 9 (03/07/14)  | Winter break                                                                                                                                                                                             |
| Week 10 (03/14/14) | Lecture: Mixed effect model I (multi-level data analysis)  
Lab: LME and GAMM functions in R – I, packages ‘nlme’, ‘lme4’ and ‘mgcv’ (HOMEWORK 3 DUE)                                                                                                          |
| Week 11 (03/21/14) | Lecture: Mixed effect model II (longitudinal data analysis)  
Lab: LME and GAMM functions in R – II                                                                                                                                                                     |
| Week 12 (03/28/14) | Lecture: Analysis of data from complex sampling design  
Lab: Package ‘survey’ in R                                                                                                                                                                               |
| Week 13 (04/04/14) | Lecture: Time-series and case-crossover analysis (by Carina Gronlund)  
Lab: Packages ‘ares’ and ‘season’ in R (HOMEWORK 4 DUE)                                                                                                                                               |
| Week 14 (04/11/14) | Lecture: Other topics and R packages not covered in this course (‘Epi’; ‘survival’; ‘NADA’)  
Lab: Review and Final Paper Q & A                                                                                                                                                                       |
| Week 15 (04/18/14) | FINAL PAPER DUE  
Lab: Review of Final Paper                                                                                                                                                                               |
**Recommended Readings**

**Week 1 (Introduction)**

**Week 4 (Effect modification)**
  - Excel spreadsheet tool to compute measures of interaction/effect modification on additive (RERI) and multiplicative scales (ratio of RRs).

**Week 5 (Statistical issues in biomarker data)**

**Week 7-8 (Smoothing)**

**Week 10 (Multilevel analysis)**

Week 11 (Longitudinal analysis)
• Park et al., Cumulative lead exposure and age-related hearing loss: The VA Normative Aging Study. Hear Res 2010;269:48-55.

Week 12 (Survey)
• Lumley. Analysis of complex survey samples. Available at http://www.jstatsoft.org/v09/a08/paper.

Week 13 (Air pollution epidemiology)
• Janes et al., Case-crossover analyses of air pollution exposure data: Referent selection strategies and their implications for bias. Epidemiology 2005;16:717-726.
Academic Integrity

From the Standard of Academic Conduct, University of Michigan School of Public Health. Found at: http://www.sph.umich.edu/students/admitted/rightsresponsibilities.html

“The faculty of the School of Public Health believes 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, 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. The preparation of reports, papers, and examinations, assigned on an individual basis, must represent each student’s own effort. Reference sources should be indicated clearly. The use of assistance from other students or aids of any kind during a written examination, except when the use of aids such as electronic devices, books or notes has been approved by an instructor, is a violation of the standard of academic conduct.”