PUB HLTH 421
INTERMEDIATE BIOSTATISTICS – 1.0 credit
Fall Quarter 2014 (September 24 – December 8, 2014)
Time: Mondays & Wednesdays 6:00-7:30 p.m.
Location: McGaw 1-401

Course instructor (office hours by appointment):
Jungwha "Julia" Lee, PhD, MPH
Assistant Professor
Department of Preventive Medicine
680 N. Lake Shore Dr., Suite 1400
(312) 503-2292
jungwha-lee@northwestern.edu (preferred)

Teaching Assistant:
Elizabeth Moore
elizabeth-yohemoore@fsm.northwestern.edu

I. Course Description
This course builds upon the material learned in Introduction to Biostatistics and Introduction to Epidemiology. Specifically, the course will focus on multivariable methods of analysis for epidemiologic and clinical studies including correlation, linear regression, logistic regression, and Cox proportional hazards regression.

• Prerequisites
  Required:
  a) Introduction to Epidemiology – PH304 (or Medical Decision Making II)
  b) Introduction to Biostatistics – PH302
  Recommended:
  a) Intermediate Epidemiology or concurrent registration in Intermediate Epidemiology – PH422
  The student MUST receive permission from the instructor if the two required courses listed above have not been completed.

II. Learning Objectives
This course is designed to familiarize students with additional statistical techniques and multivariate methods of analysis for epidemiological and clinical studies. Upon completion of this course students should be able to:

• Describe the difference between a correlation and regression analysis.
• Describe the assumptions and mechanics for estimating parameters in the simple and multiple linear regression models.
• Use Stata to perform a simple linear regression or multiple linear regression analysis, and interpret the output, including regression parameter estimates, hypothesis tests, confidence intervals, statistics which quantify the fit of the model, and potential confounding.
• Use graphical procedures in Stata for assessing whether the assumptions underlying the simple or multiple linear regression models are correct. Describe ways to remedy the problem when the assumptions appear incorrect.
• Use graphical procedures for assessing whether there are outliers in a linear regression
analysis and whether any of the observations exerts considerably more influence than the other observations in determining estimates of regression coefficients.

• Use Stata to perform an analysis of variance (ANOVA) and covariance (ANCOVA), and interpret the output, including sums of squares and F-tests.

• Use Stata for multiple comparisons and interpret results for ANOVA and ANCOVA.

• Describe the logistic regression model, its key assumptions, and their implications.

• State the relationships between odds ratios and logistic regression coefficients.

• Use Stata to fit a logistic regression model to continuous and categorical predictors and to obtain odds ratios, and interpret the output, including regression parameter estimates, hypothesis tests, confidence intervals, and statistics which quantify the fit of the model.

• Derive the odds ratio between two groups defined by their predictor values.

• Calculate and interpret the likelihood ratio test comparing two nested logistic models.

• Use Stata to obtain Kaplan-Meier survival curves, and interpret results.

• Use Stata to perform a log-rank test and interpret results.

• Define right censoring, hazard function, and proportional hazards.

• Use Stata to fit a Cox proportional hazards regression model to continuous and categorical predictors and to obtain hazard ratios, and interpret the output, including regression parameter estimates, hypothesis tests, and confidence intervals.

• Derive the hazard ratio between two groups defined by their predictor values.

• Calculate and interpret the likelihood ratio test comparing two nested Cox models.

• Detect non-proportional hazards using log-minus-log plots.

• Describe situations in which multiple linear regression analysis, logistic regression analysis, or survival analysis is needed.

• Detect and model interaction in these multi-predictor models.

• Use dummy variables to model categorical predictors and interpret results in each of these models.

• Interpret articles in the medical literature that use multiple linear regression, multiple logistic regression, or Cox proportional hazards regression.

III. Text and Readings

Required: None

If students desire additional reading on a specific topic, the following text will be on reserve in the library:


Northwestern University Library offers access to Springer eBooks.

Blackboard™

The syllabus and selected readings will be posted on the course’s Blackboard site, available at https://courses.northwestern.edu/webapps/login if you are registered for the course.

IV. Class assignments and tests

Class assignments will be posted on Blackboard. A total of 12 assignments and 2 tests (mid-term, final) will be given.
V. Student Evaluation
a) Class assignment, classroom participation – Class assignments will count 15% towards the student evaluation. Homework assigned on a Monday or Wednesday is due on the following week (except HW11). Answers to homework assignments will be uploaded on Blackboard. Missed homework assignments will not be accepted once answers have been made available. Students are expected to come to class prepared and to take part in the classroom discussions. Classroom participation and attendance will count 5% towards the student evaluation.
b) Examinations - There will be a midterm examination (30%) and a comprehensive final examination (50%). The examinations may consist of multiple choice, short answer, and statistical computations or interpretation of Stata output, and will be administered at the scheduled times. Make up examinations should be arranged in advance and will only be given under extenuating circumstances.

Letter grades will be assigned according to the following percentages:

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>93-100%</td>
</tr>
<tr>
<td>A-</td>
<td>90-92%</td>
</tr>
<tr>
<td>B+</td>
<td>88-89%</td>
</tr>
<tr>
<td>B</td>
<td>83-87%</td>
</tr>
<tr>
<td>B-</td>
<td>80-82%</td>
</tr>
<tr>
<td>C+</td>
<td>78-79%</td>
</tr>
<tr>
<td>C</td>
<td>73-77%</td>
</tr>
<tr>
<td>C-</td>
<td>70-72%</td>
</tr>
<tr>
<td>F</td>
<td>&lt;70%</td>
</tr>
</tbody>
</table>

Course evaluation
The MPH Program administers web-based course evaluations to students for each course near the end of the quarter. Your completion of both the unit (course) and faculty evaluation components is required; failure to complete either of the evaluations will result in an incomplete grade until the evaluations are submitted. You will be sent the web link and instructions via email later in the quarter. You will have about two weeks to complete the evaluations before grades are submitted.

Academic Integrity
Every Northwestern faculty member and student belongs to a community of scholars where academic integrity is a fundamental commitment. The Program in Public Health abides by the standards of academic conduct, procedures, and sanctions as set forth by The Graduate School at Northwestern University. Students and faculty are responsible for knowledge of the information provided by The Graduate School on their Web page at http://www.tgs.northwestern.edu/academics/academic-services/integrity/index.html

Academic misconduct includes, but is not limited to
1. Receiving or giving unauthorized aid on examinations or homework
2. Plagiarism
3. Fabrication
4. Falsification or manipulation of academic records
5. Aiding or abetting any of the above

The PPH follows The Graduate School’s procedure for evaluating alleged academic misconduct, as outlined on the TGS website. http://www.tgs.northwestern.edu/academics/academic-services/integrity/dishonesty/index.html

Faculty reserve the right to use the “Safe Assignment: Plagiarism Detection Tool” that is part of the Course Management System to evaluate student assignments. Information about this tool can be found at http://www.it.northwestern.edu/education/course-management/support/assessments/safeassignment.html
**VI. Class schedule**

<table>
<thead>
<tr>
<th>Week</th>
<th>Class</th>
<th>Date</th>
<th>Topic</th>
<th>Vittinghoff</th>
<th>Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>9.24</td>
<td>Wed Introduction and Overview</td>
<td>§2.1-2.5</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>9.29</td>
<td>Mon Simple Linear Regression</td>
<td>§3.3</td>
<td>HW1 due 10.06</td>
</tr>
<tr>
<td>3</td>
<td>10.01</td>
<td>Wed</td>
<td>Diagnostics in Simple Linear Regression-I</td>
<td>§4.7.1-4.7.3</td>
<td>HW2 due 10.08</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>10.06</td>
<td>Mon Introduction to Multiple Linear Regression</td>
<td>§4.7.4-4.7.6</td>
<td>HW3 due 10.13</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>10.08</td>
<td>Wed Multiple Linear Regression</td>
<td>§4.3.1-4.3.3</td>
<td>HW4 due 10.15</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>10.13</td>
<td>Mon Diagnostics in Multiple Linear Regression</td>
<td>§4.6</td>
<td>HW5 due 10.20</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>10.15</td>
<td>Wed Additional Topics in Multiple Linear Regression</td>
<td>§10.2.1-10.2.3</td>
<td>HW6 due 10.22</td>
</tr>
<tr>
<td>8</td>
<td>*8</td>
<td>10.20</td>
<td>Mon Multiple Linear Regression in the Medical Literature <em>(6-8pm)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>10.22</td>
<td>Wed</td>
<td>Analysis of Variance and Covariance</td>
<td>§3.1.4-.3.1.6</td>
<td>Exercise</td>
</tr>
<tr>
<td>10</td>
<td>10.27</td>
<td>Mon</td>
<td>Midterm examination (Covers material thru 10/20)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>10.29</td>
<td>Wed</td>
<td>Introduction to Logistic Regression</td>
<td>§5.1</td>
<td>HW7 due 11.05</td>
</tr>
<tr>
<td>12</td>
<td>11.03</td>
<td>Mon</td>
<td>Multiple Logistic Regression I</td>
<td>§5.2</td>
<td>HW8 due 11.10</td>
</tr>
<tr>
<td>13</td>
<td>11.05</td>
<td>Wed</td>
<td>Multiple Logistic Regression II</td>
<td>§5.4</td>
<td>HW9 due 11.12</td>
</tr>
<tr>
<td>14</td>
<td>11.10</td>
<td>Mon</td>
<td>Multiple Logistic Regression in the Medical Literature <em>(6-8pm)</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>11.12</td>
<td>Wed</td>
<td>Introduction to Person-Time Data</td>
<td>§3.5</td>
<td>HW10 due 11.19</td>
</tr>
<tr>
<td>16</td>
<td>11.17</td>
<td>Mon</td>
<td>NO CLASS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>11.19</td>
<td>Wed</td>
<td>Introduction to Cox Proportional Hazards Regression</td>
<td>§6.1</td>
<td>HW11 due 11.24</td>
</tr>
<tr>
<td>18</td>
<td>11.24</td>
<td>Mon</td>
<td>Cox Proportional Hazards Regression</td>
<td>§6.2, 6.4</td>
<td>HW12 due 12.01</td>
</tr>
<tr>
<td>19</td>
<td>12.01</td>
<td>Mon</td>
<td>NO CLASS (THANKSGIVING 11.27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>12.03</td>
<td>Wed</td>
<td>Course Review</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>12.08</td>
<td>Mon</td>
<td>Comprehensive Final Examination</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Note Oct. 8th, Nov. 10th, and Dec. 1st will be a 2-hour class (6-8pm).

**See Blackboard for class material.

**X. Additional resources**

Homework assignments require that students do analyses using STATA/IC 12. STATA is available for student use in the Galter Library Computer Center. If a student wishes to purchase STATA to install on his/her own computer, Northwestern participates in STATA’s “direct ship grad plan” pricing ([http://stata.com/order/new/edu/gradplans/gp-direct.html](http://stata.com/order/new/edu/gradplans/gp-direct.html)), and the student can order directly from the STATA website. See PPH Student Handbook to obtain own software license(s).

Rev. 6.4.2014