Biost 518 / Biost 515
Applied Biostatistics II / Biostatistics II

Syllabus
Winter 2015

Note: *Students are responsible for knowing all information provided on this syllabus. This syllabus is accurate as of the beginning of the course. Students are further responsible for any changes to this information as announced in class, posted on the web pages, or announced via email.*

Instructor : Scott S. Emerson, M.D., Ph.D., Professor of Biostatistics
Office : HSB H655-J
Phone : 616-6678
Email : semerson@uw.edu
Office hours : W 10:30a - 1:30 or by appointment

Assistants : LaNae Schaal (lschaal@uw.edu)
Office hours : Su 3:00p - 5:00p HSLIC
Tu 1:00p - 3:00p HSLIC
Th 8:00a - 10:00a HSLIC

Jon Fintzi (fintzij@uw.edu)
Office hours : We 2:30p - 4:00p HSLIC

Time and Place : Lectures : MWF 9:30a - 10:20a HSB T747
Disc AA M 8:30a - 9:20a HSB T478
Disc AB W 8:30a - 9:20a HSB T478
Disc AC F 8:30a - 9:20a HSB T531

Class Web Pages: `http://www.emersonstatistics.com/b518/`

The web page will be used to post datasets, notices, handouts, etc. I urge you to check this site regularly. Questions that are submitted to me (via email or otherwise) that I think might be of general interest will have their answers posted on the web page, as well.

Prerequisites : Biost 517 / Biost 515 or equivalent

Optional Text : Vittinghoff, Glidden, Shiboski, McCulloch
Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, 2nd edition. Springer.

Other Texts : Kleinbaum, Kupper, Muller, and Nizam (KKM)
Applied Regression Analysis and Other Multivariable Methods, 3rd ed.
Duxbury Press, 1998
Kleinbaum, *Logistic Regression: A Self-Learning Text*, Springer, 1994
Kleinbaum, *Survival Analysis: A Self-Learning Text*, Springer, 1996

Note : These texts are optional, because I believe they are often not teaching the robustness of regression methods. When the texts disagree with lecture material, lecture material takes precedence.

Computing : Software : R or Stata

Weekly homeworks will involve statistical analyses that will generally require access to statistical software. While students may most often use the statistical software of their choice, so long as the software is capable of performing the necessary statistical procedures, access to data sets and instructions for performing analyses with computing assumes the use of R or Stata. R is a freely available software package for which we have developed a suite of routines that perform the analyses typically required for this class. Stata is available on the computers in the HSLIC. Instructions for obtaining personal copies of R or Stata are available on the class website.

Attendance : Lectures : Generally required (for quizzes)
Discussions : Generally required for contribution to the discussion

Assignments : Written problem sets approximately weekly
Weekly data analyses for discussion sections
Written quizzes and discussion during lecture on an occasional basis
One midterm (in class, closed notes and requires a calculator that cannot access the internet)

.

One written report of a data analysis
Final exam (in class, closed notes)

Homework problems requiring a written solution will be due approximately weekly. These assignments will consist of applications of statistical methods to real data analyses. Students in Biost 515 may have additional problems assigned. These problems will tend to focus on the theory underlying the methods covered in class. These latter problems can be worked by enterprising Biost 517 students as extra credit.

Students are encouraged to seek help from the instructor, the TAs or other students with the written homework problems. However, the work that is handed in should reflect only that student's work. That is, obtaining help from other students in order to learn the METHODS of solution is allowed, but copying another student's answer is NOT. Assignments handed in late will not be accepted unless pre-approved. We reserve the right to grade only selected portions of the written homework. The weekly data analyses for discussion and the written reports are described under the Data Analysis Laboratory.

Homeworks will be peer graded. Students will be expected to submit their homeworks electronically using an anonymized code. **It is the student's responsibility to ensure that no personally identifying information is included in their submitted homework or grading.** Students will then be assigned to grade another student's homework guided by a key and grading instructions provided by the instructor or TAs. A random sample of papers will also be graded by the TAs. Any student may also appeal to the instructor regarding any grade assigned by a peer.

Quite often, a small portion of class will be devoted to a short quiz. Students will be asked to prepared to 1) answer brief written questions about preparatory material for that lecture at the beginning of class, and 2) discuss their answers to the questions during class.

On one occasion during the quarter, the instructor will designate a data analysis requiring a written report from the students. The length of the report should be approximately 10-12 pages, and it should be written to a statistically naive reader. This will be a group project (except for the Statistical Analysis Plan), and reports will be "refereed" by other groups. Further details (and examples) will be distributed later in the quarter.

Data Analysis Laboratory:

The discussion sections will be used as a data analysis laboratory in which it is envisioned that the students will gain experience in the general approach to a data analysis and in the application of the statistical methods learned in lecture. Each week, a data analysis problem will be assigned. Students will be expected to analyze the data set to address the question of interest and to come to the discussion section prepared to answer questions about their methods and results. Students will often be called on at random. Because this is a learning situation, it is not expected that a student will necessarily have an error-free analysis. It is expected that a student will spend 1-2 hours each week thinking carefully about the problem and attempting to apply good statistical principles to its solution. A student's contribution to the discussion of the data analysis problem will be evaluated on the following scale: 0= clearly inadequate attention to the assignment, 8= analysis attempted but major problems with the approach, 9= analysis attempted and some thought given to the best approach, 10= exemplary insight into the problems posed by the data analysis.

Grading	:	Written homeworks	20%
		Quizzes and oral discussion	10%
		Midterm	25%
		Report	20%
		Final examination	25%

Additional Resources

1. The following materials will be posted on the webpages:
   a. Copies of the PowerPoint slides used in lectures. The dates for each lecture are approximate, and a given lecture period may cover material from more than one handout.
   b. Supplemental notes that will not be covered in lecture, but may be of use in preparing for the data analysis laboratory.
   c. Supplemental notes on material that should be a review for most students, but which some students may need to study in detail. This material will not be covered in class.
   d. Homeworks, exams, and keys from previous quarters that I taught this class.
   e. Homework assignments (typically posted on Wednesdays and due the following Wednesday).
   f. Keys to homeworks and exams from this quarter (only after the due date).
   g. Streaming video of the computer display and audio of the lecture.
2. Electronic mail (e-mail) will be used for communication of errata and other announcements that are of interest to the general class. All students are required to ensure that they receive emails sent to the address of record with the UW Registrar. Throughout the quarter, students may submit questions regarding the course material via e-mail. Answers to questions that I feel are of general interest will be broadcast to the entire class (the identity of the source of the question will be protected). Questions that are likely to be of interest only to a single student will usually be answered individually. I try for reasonably quick turnaround on email questions, but backlogs do occur. It may happen that I think I have answered your question in a general message broadcast to the class, but you are still unsure of the answer. Do not hesitate to send your question again, and I will try to address it further.

Course Objectives

This course builds on the material of Biost 517/ 514, and thus it is assumed that you are familiar with the basic principles of descriptive and inferential statistics related to means, proportions, and survival estimates in one and two sample problems and the descriptive (and some inferential) statistics related to simple linear regression and correlation. It is also assumed that you have some familiarity with the most basic aspects of confounding and effect modification. At the end of Biost 518 / 515, you should be able to:

1. Identify the type of question for which a regression analysis might be appropriate.
2. Perform suitable descriptive analyses of the data.
3. Develop a regression model and perform an analysis using statistical software, including
    a. definition of dependent and independent variables,
    b. appropriate choice of summary measure for modeling,
    c. appropriate choice of transformations,
    d. use of dummy variables where indicated,
    e. appropriate selection of variables to include in the model, and
    f. correct modeling of interactions as necessary.
4. Perform an analysis of variance, including proper adjustment for multiple comparisons when evaluating mean contrasts.
5. State the statistical assumptions that are the basis for the conclusions of your analysis.
6. Present the results of your analysis to a statistically naive reader, including a full interpretation of all parameter estimates.

In this course I will stress the philosophy and principles behind the statistical methods rather than the formulas used to implement the methods. The course is targeted to students who want to be able to read biomedical research literature critically, as well as to students who will eventually be analyzing data as a part of their research. Even so, I expect that most students will find the course demanding and some of the material difficult. I welcome student suggestions regarding ways in which this goal can be best achieved. If you have questions regarding the content or structure of the class, please feel free to talk (or write) to me at any time during the quarter.

Biost 518 / Biost 515 Course Outline
Winter 2015

The following is a tentative outline of the topics to be covered during the quarter. We reserve the right to modify this outline as conditions require.

Lect	Date	Day	Topic	Hand In	
1.	5 Jan	Mon	Organization, Review of two sample problem		
2.	7 Jan	Wed	General regression model; inferential assumptions		
3.	9 Jan	Fri	Simple linear regression		
4.	12 Jan	Mon	Simple linear regression: extensions	HW #1	
5.	14 Jan	Wed	Simple logistic, Poisson regression		
6.	16 Jan	Fri	Simple proportional hazards regression	Grading: HW #1	
	19 Jan	Mon	Holiday - No class	HW #2	
7.	21 Jan	Wed	Transformation of predictors: univariate		
8.	23 Jan	Fri	Transformation of predictors: dummy variables	Grading: HW #2	
9.	26 Jan	Mon	Transformation of predictors: splines	HW #3	
10.	28 Jan	Wed	Precision of inference: sample size computation		
. 11.	30 Jan	Fri	Adjustment for covariates	Grading of HW #3	
12.	2 Feb	Mon	Stratified analyses	HW #4	
13.	4 Feb	Wed	Confounders		
14.	6 Feb	Fri	Precision variables	Grading: HW #4	
15.	9 Feb	Mon	Effect modification	HW #5	
16.	11 Feb	Wed	Interactions		
	13 Feb	Fri	**MIDTERM** (in class, closed book, closed notes)	Grading: HW #5	
	16 Feb	Mon	Holiday - No class	HW #6	
17.	18 Feb	Wed	Clustered data		Project SAP
18.	20 Feb	Fri	Weighted regression	Grading: HW #6	
19.	23 Feb	Mon	Time varying covariates	HW #7	
20.	25 Feb	Wed	Model building: Selection of variables		
. 21.	27 Feb	Fri	Model building: Form of variables	Grading: HW #7	
22.	2 Mar	Mon	Model and case diagnostics	HW #8	
23.	4 Mar	Wed	Prediction		Draft Project
24.	6 Mar	Fri	Prediction	Grading: HW #8	Referee Report
25.	9 Mar	Mon	Highly predictive models	HW #9	
26.	11 Mar	Wed	Missing data		
27.	13 Mar	Fri	Overview and Review	Grading: HW #9	Final Project
	18 Mar	Wed	**FINAL EXAM** 8:30 am - 10:20 am (in class, closed book, closed notes)	Final Exam	