

**Biost 518**  
**Applied Biostatistics II**

**Syllabus**  
Winter 2008

**Instructor** : Scott S. Emerson, M.D., Ph.D., Professor of Biostatistics  
Office : HSB F673  
Phone : 543-1044 (Biostatistics)  
          221-4185 (ROC)  
Email : semerson@u.washington.edu  
Office hours : immediately after class or by appointment

**Assistants** : Daryl Morris (darylm@u.washington.edu)  
Office hours : Tu 6:00p - 7:30p HSLIC  
              Th 1:30p - 3:00p HSLIC  
  
Rob Wellman (wellmanr@u.washington.edu)  
Office hours : MF 12:30p - 2:00p HSLIC

**Time and Place** : Lectures : MWF 9:30a - 10:20a HSB T439  
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 T474

**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 or equivalent

**Text** : 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

**Computing** : Software : 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 help with computing assumes the use of Stata. Stata is available on the computers in the HSLIC. Instructions for obtaining personal copies of Stata are available on the class website.

**Attendance** : Lectures : Highly recommended  
Discussions : Highly recommended

<b>Assignments</b>	:	Written problem sets approximately weekly Weekly data analyses for discussion sections One midterm (in class, closed notes) 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 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.

#### **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. 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 2-3 hours each week thinking carefully about the problem and attempting to apply good statistical principles to its solution.

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, and reports will be "refereed" by other groups. Further details (and examples) will be distributed later in the quarter.

<b>Grading</b>	:	Written homeworks      25%
		Midterm                  25%
		Report                  20%
		Final examination      30%

#### **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 submit their e-mail address within the first two days of class. 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.

3. I have requested that a number of texts on statistical methods (especially regression methods) be placed on reserve at the Health Sciences Library.

### **Course Objectives**

This course builds on the material of Biost 517, 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 inferential statistics related to simple linear regression and correlation. It is also assumed that you are familiar with the most basic aspects of confounding and effect modification. At the end of Biost 518, 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 Course Outline  
Winter 2008

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.

	Date	Day	Topic	Hand In
1.	7 Jan	Mon	Organization, Setting	
2.	9 Jan	Wed	Precision of Inference	
3.	11 Jan	Fri	Confounding, Effect Modification	
4.	14 Jan	Mon	Stratified analyses	
5.	16 Jan	Wed	Review of simple linear regression	HW #1
6.	18 Jan	Fri	Simple logistic, Prop Hzd regr	
	21 Jan	Mon	HOLIDAY - No class	
7.	23 Jan	Wed	Simple Poisson regression	HW #2
8.	25 Jan	Fri	Adjustment for Covariates	
9.	28 Jan	Mon	Inference in multiple regression	
10.	30 Jan	Wed	Effect modification: binary predictors	HW #3
11.	1 Feb	Fri	Effect modification: cts predictors	
12.	4 Feb	Mon	Modeling dose response; ANOVA	
13.	6 Feb	Wed	Choice of predictor transformation	HW #4
14.	8 Feb	Fri	Flexible modeling of predictors	
15.	11 Feb	Mon	MIDTERM (in class)	Midterm
16.	13 Feb	Wed	Case study	
17.	15 Feb	Fri	Clustered observations	
	18 Feb	Mon	HOLIDAY - No class	
18.	20 Feb	Wed	Clustered observations	HW #6
19.	22 Feb	Fri	Time varying covariates	
20.	25 Feb	Mon	Weighted regression	
21.	27 Feb	Wed	Prediction of summary measures	HW #7
22.	29 Feb	Fri	Prediction of individual measurements	
23.	3 Mar	Mon	Model diagnostics	
24.	5 Mar	Wed	Case diagnostics	HW #8
25.	7 Mar	Fri	Missing data	Report (init)
26.	10 Mar	Mon	Exploratory models	
27.	12 Mar	Wed	Exploratory models	HW #9
28.	14 Mar	Fri	Overview and Review	Referee
	19 Mar	Wed	FINAL EXAM 8:30 am - 10:20 am	Final Exam