Biost 517 / Biost 514
Applied Biostatistics I / Biostatistics I

Syllabus
Fall, 2011

Instructor : Scott S. Emerson, M.D., Ph.D., Professor of Biostatistics
Office : HSB F673
Phone : 616-6678 (Biostatistics)
Email : semerson@uw.edu
Office hours : Tu 2:00p - 5:00p
(or by appointment)

Assistants : GregLevin (glevin11@uw.edu)
Office hours : M 12:30p - 2:00p HSLIC
: F 12:30p - 2:00p HSLIC

Elisa Sheng (esheng@uw.edu)
Office hours : M 2:00p - 3:30p HSLIC
: F 2:00p - 3:30p HSLIC

Time and Place :

Lectures	MWF	9:30a - 10:20a	HST 439
Disc AA	M	8:30a - 9:20a	HST 473
Disc AB	W	8:30a - 9:20a	HST 473
Disc AC	Tu	8:30a - 9:20a	HST 478
Disc AA	F	8:30a - 9:20a	HST 473

Class Web Pages: `http://www.emersonstatistics.com/b517/`

The web page will be used to post copies of the PowerPoint slides presented in lecture, homework assignments, datasets, documentation, other information, etc. I urge you to check this site regularly.

Prerequisites : None

Optional Text : Rosner: *Fundamentals of Biostatistics*, 7th ed., Duxbury, 2000

Computing : Software : Stata (or, for the brave, R)

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, help with computing assumes the use of Stata or R. Stata is available on the computers in the HSLIC. Instructions for obtaining personal copies of Stata are available on the class website. R is freely available on the web, but is generally more difficult to learn and use. Computing commands covered in lecture will be Stata, though we are happy to help with R in office hours.

Attendance : Lectures : Occasionally required, always highly recommended
Discussions : Required

Assignments : Written problem sets approximately weekly
Weekly data analyses for discussion sections
Written quizzes and discussion during lecture approximately weekly
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. Biost 514 may occasionally have an extra, more theoretical or computationally intensive problem assigned. 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.

Approximately once per week, students will be asked to view a lecture on line and be prepared to 1) answer brief written questions about that lecture at the beginning of class, and 2) discuss their answers to the questions during class.

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 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.

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 5 - 7 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.

Grading :

Written homeworks	20%
Quizzes and oral discussion	10%
Midterm	25%
Report	15%
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 (most times and only after the due date).
   g. (Technology and skill willing) Streaming audio/video of lectures and discussion sections.
2. Electronic mail (email) will be used for communication of errata and other announcements that are of interest to the general class. I will use the email address supplied by the university course registration list. It is the student's responsibility to ensure that they are receiving emails at their desired email address. Throughout the quarter, students may submit questions regarding the course material via email. 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 elementary texts on statistical methods be placed on reserve at the Health Sciences Library. These include
   a. Rosner B, *Fundamentals of Biostatistics, 5th ed.* (QH323.5.R822)
   b. Pagano M and Gauvreau, *Principles of Biostatistics* (QH323.5.P34)
   c. Fisher and van Belle, *Biostatistics: A Methodology for the Health Sciences* (QH323.5.F57)
   d. Moore DS and McCabe GP: *Introduction to the Practice of Statistics*, 3rd ed. (QA276.12.M65)
   e. Fleiss J, *Statistical Methods for Rates and Proportions*
   f. Kleinbaum DG, *Survival Analysis: A Self-Learning Text* (WA960.K64)
   g. Parmar MKB and Machin D, *Survival Analysis: A Practical Approach* (WA950.P254s)
   h. Klein JP and Moeschberger ML, Survival Analysis: Techniques for Censored and Truncated Data (WA950.K63s)

Course Objectives

This course provides an introduction to the statistical analysis of data. Emphasis is placed on the analysis of data to answer scientific questions. Thus the major objectives of this course are

1. To explore the ways in which statistical methods can be used to address scientific questions,
2. To present simple data analysis methods, and
3. To teach a general approach to a data analysis problem.

To those ends, this course will stress the general abstraction of descriptive and inferential statistics to address a scientific question. We will primarily address methods for the setting of one response variable and one grouping variable. This includes one and two sample problems, one way analysis of variance, and simple regression. Late in the quarter we will address stratified analyses. (Biost 518/515 will cover multivariable regression.) Topics covered will include definition of common descriptive techniques, estimation and testing for continuous, discrete, and censored response variables in parametric models, and semiparametric and nonparametric alternatives to those tests, including Monte Carlo methods. Emphasis will be placed on the similarity among the various forms of analyses.

At the end of Biost 517/514, a student should have made significant progress toward being able to:

1. Demonstrate an organized approach to the analysis of data gathered to address a scientific question.
2. Develop an appropriate statistical model to analyze such data to address a scientific question, including
   a. refinement of vaguely stated scientific hypothesis into a statistical framework,
   b. identification of the dependent (response) variable, including a reasonable probability model for that response and a summary measure to be estimated and/or tested,
   c. identification of the independent (predictor) variables denoting any groups to be compared.
3. Perform suitable descriptive analyses of the data, including descriptions relevant to the sampling scheme and any patterns of missing data.
4. Compute estimates and/or test statistics using standard statistical software.
5. Make statistical inference about the generalizability of the analysis results to a larger population.
6. State any statistical assumptions that are the basis for the conclusions of your analysis.
7. Perform analyses to determine whether the assumptions are sensible both on sample-wide and individual case bases.
8. Present the results of your analysis to a statistically naive reader, including a full interpretation of all parameter estimates.

Biost 517 / 514 Course Outline
Fall 2011

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. ("Rosner" refers to relevant sections in the textbook by Rosner.)

	Date	Day	Topic	Rosner	Hand In
1.	23 Sep	Wed	Course organization, Overview	1	
2.	30 Sep	Fri	Scientific questions answered with statistical analyses of data		
3.	3 Oct	Mon	Descriptive statistics overview: purpose, classification of variables	2.1	
4.	5 Oct	Wed	Descriptive statistics overview; location		HW #1
5.	7 Oct	Fri	Univariate description of location: means, others	2.2-3	
6.	10 Oct	Mon	Univariate description: spread, skewness	2.4-11	
7.	12 Oct	Wed	Univariate description: entire distn		HW #2
8.	14 Oct	Fri	Censored data	14.8-9	
9.	17 Oct	Mon	Censored data		
10.	19 Oct	Wed	Bivariate descriptive statistics	11.1-3	HW #3
11.	21 Oct	Fri	Correlation	11.7	
12.	24 Oct	Mon	Introduction to inference	6.1-2;7.1-2	
13.	26 Oct	Wed	Sensitivity, specificity, Bayes rule		HW #4
14.	28 Oct	Fri	Probability	3;4;5	
15.	31 Oct	Mon	One-sample inference for means	7.3-4	
16.	2 Nov	Wed	Generalizations of one-sample inference	8.2-3	HW #5
17.	4 Nov	Fri	MIDTERM (in class, closed book)		MIDTERM
18.	7 Nov	Mon	One-sample inference for proportions (incl censored data)		
19.	9 Nov	Wed	Two-sample inference for means	8.4-5;8.7	
	11 Nov	Fri	HOLIDAY (no class)		
20.	14 Nov	Mon	Generalizations of two-sample inference	9.2-4	HW #6
21.	16 Nov	Wed	Two-sample inference for binary data		
22.	18 Nov	Fri	Two-sample inference with censored data	14.10	HW #7
23.	21 Nov	Mon	Dependent data within clusters		
24.	23 Nov	Wed	Dependent data within clusters		
	25 Nov	Fri	HOLIDAY (no class)		
25.	28 Nov	Mon	One-way ANOVA / Simple linear regression	11	HW #8
26.	30 Nov	Wed	Simple linear regression		
27.	2 Dec	Fri	Simple linear regression: Transformations		Draft Report
28.	5 Dec	Mon	Simple logistic regression		
29.	7 Dec	Wed	Simple logistic / PH regression		Referee Report
30.	9 Dec	Fri	Simple PH regression		HW #9
	14 Dec	Wed	FINAL EXAM 8:30 am - 10:20 am		Final Exam / Report