**Biost 517**
**Applied Biostatistics I**

**Syllabus**
Fall, 2006

**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 : Tu 9:00a - 10:00a
F 10:30a - 11:30a (or by appointment)

**Assistants** : Mark Giganti (giganti@u.washington.edu)
Office hours : M 1:00p - 3:00p HSLIC
Tu 5:00p - 7:00p HSLIC
F 1:30p - 3:30p HSLIC

: Jessie Gu (wengu@u.washington.edu)
Office hours : Tu 2:00p - 5:00p HSLIC
W 1:00p - 4:00p HSLIC

**Time and Place** : Lectures : MWF 9:30a - 10:20a HSB D209
Disc AA M 8:30a - 9:20a HSB T530
Disc AB W 8:30a - 9:20a HSB T473
Disc AC F 8:30a - 9:20a HSB T531

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

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** : None

**Texts:**
Required : Rosner: *Fundamentals of Biostatistics*, 5th ed., Duxbury, 2000

**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 (but videotaped)
Discussions : Highly recommended (and not videotaped)

**Assignments** : 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 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 | 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).
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 unless informed otherwise. 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 asked for lectures to be videotaped, with the tapes available at the Reserve Desk in the Health Sciences Library.

4. 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 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, 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. Perform suitable descriptive analyses of the data.
3. 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.
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 Course Outline
Fall 2006

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