BIOS 501, Statistical Methods II

INSTRUCTORS: Paul Weiss
GCR 308
Phone: (404) 712-9641
FAX: (404) 727-1370
e-mail: pweiss2@sph.emory.edu

LECTURES: T TH 8:30 - 9:50 pm Location: CNR 1000

PREREQUISITES: BIOS 500, or permission of instructor

LABS:

Day/Time	
T	4:00 - 5:50 pm
T	6:00 - 7:50 pm

OFFICE HOURS: Paul Weiss: Tuesdays 10 am - 12 pm (otherwise by appointment if necessary)
Please e-mail at least 24 hours in advance in order to schedule a time with me. If a face-to-face meeting is not essential, ask questions on the class web site discussion board in order to get a faster answer (often in ≤ 24 hours.)
Teaching Assistants: See Class web site for hours.

TEXTBOOKS:

Recommended: Course text from Bios 500 (as a reference)

On Reserve: Weiss, N.A. (1995). *Introductory Statistics* (8th ed.). Addison-Wesley Publishing Company.

Daniel, W.W. (2005) *Biostatistics* (8th ed.). Wiley.

Rosner, B.A. (1995) *Fundamentals of Biostatistics* (5th ed.). Belmont: Wadsworth Publishing Company.

D.G. Kleinbaum, L.L. Kupper, and K.E. Muller, A Nizam. *Applied Regression Analysis and Multivariable Methods*, 3rd edition, Duxbury Press, 1998.

I will also be posting chapters of a book I am writing on materials covered in this class. I am hoping that, with feedback provided from the students, I can formalize them and get the book published soon. These chapters will cover the course material more closely than will any of the above texts.

COURSE DESCRIPTION

This course is the follow-up to Biostatistical Methods I. Students will apply many of the concepts learned in the first semester in a broader field of statistical analysis: model construction. Topics covered include Linear Regression, Analysis of Variance, Logistic Regression and Survival Analysis. Students who successfully complete this course will have a deep understanding of many analytical methods used by public health researchers in daily life.

OBJECTIVES Upon completion of the course, students will be able to:

1. Choose appropriate methods, models, parameters and hypotheses for a variety of problems related to simple and multiple linear regression, logistic regression and survival analysis.
2. Run SAS programs and interpret SAS output from these programs.
3. Check assumptions underlying the regression methods.
4. Use related statistical tables.

GRADING SCALE

A:	[96-100+	A-:	[91-96)	
B+:	[86- 91)	B:	[81-86)	B-: [76-81)
C:	[66- 76)			
F:	Below 66			

EVALUATION

Homework	10%
Exam I:	20%
Midterm Paper:	35%
Final Exam:	35%

WEB SITE **http://blackboard.sph.emory.edu/**
The web site will be our primary means of communication outside of class.

- Instructors will post announcements for lecture and lab on the web site.
- You may post questions, and lab instructors or TAs will then post answers, on the discussion boards located on the class web pages.
- **Lecture sessions will be recorded regularly and links to the recordings will be posted on the website. These recordings will also contain a scan of the document camera slides produced in class. Students have open access to the recordings and may watch them as many times as they would like.**
- Homework assignments will also be posted on the web as well as solutions to those assignments.

- **You are expected to access the web site daily, M-F, to look for newly posted documents and discussion board messages. It is suggested that you also check the site on weekends, for updates to posted documents and other important announcements.**

General Notes

- Please attend all lectures and labs in order to avoid falling behind. It will be difficult to fully understand new material being presented in class and labs if you fall behind. We realize that many of you are working full or part-time, and that your jobs require you to travel out of town occasionally. Please arrange your schedule in such a way that class and lab absences are kept to a minimum. If you do miss a class or lab, it is your responsibility to catch up, and you should do so immediately.

- Use the web-page discussion boards and specified office hours freely. As soon as a course-related question arises, post it on the appropriate web site discussion board. If the question is more private in nature, e-mail the instructor directly (pweiss2@sph.emory.edu for class-related or lab-related questions), or get in-person help from the instructor or a TA during the next available office hour. Ask questions far in advance of quizzes and exams!

Homework

- The homework for this class will be helpful in preparing you for the midterm paper and exams. The first homework set and homework sets assigned after the midterm paper will be graded by a short homework quiz comprising ten true/false questions. The three homework sets prior to the midterm paper will be collected and assessed for completeness but will not be graded for accuracy. These three assignments are designed to facilitate your working with the midterm data and prevent you from waiting until the last minute to get working on your paper.

- It is highly recommended that you attempt all of the homework provided in class. These homework sets will serve as guidelines in helping you prepare for the exams, and also help you in formulating analyses germane to the midterm paper. Many of the homework sets will feature open-ended questions, designed to get you to think analytically, and may have a variety of correct answers.

- Working together on homework problems is acceptable but I strongly suggest you make a solid attempt at the work prior to seeking out help from your friends, TAs or instructors. Students who make a real effort on their own tend to fare better in the class compared to students who seek help without trying to figure things out on their own first.

- The lab is an integral part of this course, but your performance in lab will not have any impact on your class grade. Techniques we cover in class will be reinforced in lab, and methods you learn in lab will help you in completing homework assignments and projects. Keeping up with lab assignments will improve your chance of success in class.

Quizzes and Exams

- This class is much more computationally intensive than was BIOS 500. To this end, in-class exams could be much more cumbersome to complete in the time allotted. I have divided the term into three sections: Simple Linear Regression, Multiple Linear Regression and Model Selection, and Logistic Regression and Survival Analysis. Upon completion of the first section, an in-class exam will test students on the theory and application of simple linear models. The second section will culminate in a paper where the student will present a methodology for using multiple linear regression to answer questions about a data set. The final exam will have a take-home component as well as an in-class component.

- In-class exams will be open book and open note. A calculator is a good idea but not a necessity. You may use your smartphone as a calculator as long as you are not using it as a communication device to elicit answers from an outside party during the exams. Using your phone to receive or provide test answers during an exam is an honor code violation and could result in a F grade in the course and/or expulsion from the school if you are caught.

- Exams will be held in class on the days specified on pages 5-6 on this syllabus (dates are tentative). I try to arrange extra time for the final exam so that time pressure will not be a factor. If you have a course-related or unavoidable personal conflict with the exam time, please e-mail the instructor at least ten days before the exam -- another exam time for the same day will then be arranged.

- ***Please do not schedule out-of-town trips or other appointments on exam and quiz dates. In case of emergency or unavoidable conflicts, I will, of course, do my best to accommodate you. In all other cases make-up exams will not be allowed.***

- Prior to exams, you may be required to conduct data analyses using SAS. The output that you produce may be needed to answer questions during the in-class exams.

- Failure to take the final exam or hand in a midterm paper will result in a course grade of 'F'.

- Review sessions will likely be held before each exam. In the event of school closings as a result of inclement weather it may be necessary to forego a review session to maintain the schedule of topics.

- **Students are <u>NOT</u> allowed to work together on exams, or on any work that is related to quizzes or exams**. **All quiz and exam-related work must be done completely independently and must not be discussed with anyone except Paul Weiss, the lab instructors or the teaching assistants.**

- The midterm paper will be a data analysis project, due after the mid-semester break. This paper will not exceed three (3) pages, including graphs, tables and supplementary material. The paper will be due on **MARCH 22** (this is the second Tuesday after Spring Break) at the **BEGINNING OF CLASS**. Please note that **NO LATE PAPERS WILL BE ACCEPTED** and if you do not turn in your paper on time you will receive a zero for the paper and a failing grade for the semester. Family emergencies, printer failures, oversleeping and incarceration are **NOT** acceptable excuses for handing the paper in late. If you are unable to come to class on time you may e-mail your paper to me prior to the start of class. Papers received prior to the start of class via e-mail (with a time stamp indicating time prior to 8:30 am) will be accepted as on-time. More information on this project will be provided after the simple linear regression quiz.

  There is a blackboard item where students are encouraged to post their midterm papers. This item is set up to allow multiple postings; students may post their paper any time prior to the due date and time. Students submitting multiple attempts will have their last attempt graded. The link to submit the paper will disappear at 8:30 am on the date the paper is due. **No papers will be accepted after class begins on March 22.** If you would like to submit a hard-copy (in addition to or in lieu of the electronic copy) you may hand it in at the beginning of class on the day it is due.

BIOS 501: TENTATIVE LECTURE OUTLINE

Date	Topics	Reading
1/12	Overview of course, introduction to covariance and correlation	W[2]: 14.4 R[1]: 11.1, 11.7-11.8 D[3]: 9.1, 9.6 – 9.8 DA[4]: 10.1
1/14	Finish covariance and correlation, start simple linear model	W: 14.4 R: 11.1, 11.7-11.8 D: 9.1, 9.6 – 9.8 DA: 10.1
1/19	Simple Linear regression: general concepts, method of least squares, hypothesis tests on parameter estimates	W: 14.1 – 14.3, 15.2 R: 11.2 – 11.4 D: 9.2, 9.3
1/21	ANOVA table for SLR: F-test inferences on parameters and the coefficient of determination	D: 9.4, 9.5 R: 11.4
1/26	Confidence intervals for parameter estimates and predicted values	W: 15.2 – 15.4
1/28	Multiple regression: Introduction, general concepts, hypothesis tests	W: Module A (CD-ROM in book)
2/2	Catch-up and Review	
2/4	**Exam I: Simple Linear Regression. IN CLASS**	
2/9	ANOVA for Multiple Regression: F-test inferences; the extra sum-of-squares principle; interactions	R: 11.9 D: Chapter 10
2/11	Multiple and Partial F-tests for multiple regression	R: 11.9 D: Chapter 10
2/16	Comparing regression results for two or more groups: dummy variables, coding schemes.	D: Chapter 10 R: 11.8
2/18	Selecting the best regression equation: stepwise regression	K[3]: 16.1 - 16.6 W: Module B; B5
2/23	Selecting the best regression model: All-possible-regressions	K: 16.1-16.6

1. R: Rosner
2. W: Weiss
3. K: Kleinbaum, et al.

BIOS 501: TENTATIVE LECTURE OUTLINE (Continued)

Date	Topics	Reading
2/25	Building Associative Models	
3/1	Checking the assumptions for regression (Regression Diagnostics)	W: 15.1, 15.5
3/3	Catch-up, Review, and Questions for Midterm Project	W: 15.1, 15.5 W: Module B; B6
3/7 – 3/11	**Spring Break** (No classes)	
3/15	Maximum Likelihood Estimation	
3/17	Logistic Regression: Review of 2x2 tables; odds ratios;	K: 22, 23.1-23.4 R: 13.7 D: 11.4 DA: 11.1-11.7
3/22	Logistic Regression: maximum likelihood estimation; logistic regression model (SAS Output) **Paper due at beginning of class. No late papers will be accepted.**	K: 22, 23.1-23.4 R: 13.7 D: 11.4 DA: 11.1-11.7
3/24	Logistic regression: odds and odds ratio (OR) calculations	K: 23.1 – 23.4
3/29	Logistic regression: analysis strategy, CIs for odds ratios, prediction and interpretation issues.	
3/31	Interactions; interpreting odds ratios in the presence of interactions	
4/5	Validating predictive logistic regression models	
4/7	Applications of Logistic Regression: Propensity Scores	
4/12	Logistic Regression Model Residuals and Model Fit Statistics	

BIOS 501: TENTATIVE LECTURE OUTLINE (Continued)

Date	Topics	Reading
4/14	Survival Analysis: introduction, Kaplan-Meier estimator	R: 14.8-14.9 D: 12.8
4/19	Survival Analysis: proportional hazards regression models	
4/21	Review	
4/26	**Exam III. Focus on material since Midterm Paper**	