

EPIDEMIOLOGIC ANALYSIS OF BINARY DATA

Epidemiology 718 (269)

Fall Semester, 2006

Instructors:

Jay S. Kaufman, Ph.D
email: Jay_Kaufman@unc.edu
Department of Epidemiology
UNC School of Public Health
2104C McGavran-Greenberg Hall
phone: 919-966-7435
fax: 919-966-2089

David B. Richardson, Ph.D
Email: David_Richardson@unc.edu
Department of Epidemiology
UNC School of Public Health
2102B McGavran-Greenberg
Phone: 919-966-2675
Fax: 919-966-2089

Teaching Assistants:

Lisa C Vinikoor [vinikoor@email.unc.edu] [Stata]
Sarah J Nyante [nyante@email.unc.edu]
Nanette R Lee [nlee@email.unc.edu] [Nutrition]

Time and Place:

Wednesday	11:00-11:50, MC 1305
Friday	11:00-12:50, MC1305

Usual Course Structure:

Wednesday	11:00-11:50	Lecture
Friday	11:00-11:50	Lecture
Friday	12:00-12:50	Lab

Lectures:

Lectures are didactic, but your active participation (in the form of questions) is encouraged, as long as it doesn't lead to extensive digressions.

Lab:

Lab often contains a didactic component, but you are strongly encouraged to use this time to ask practical questions about any confusion or difficulty you face in the practical implementation of the course material. Because lectures need to move forward at a somewhat unrelenting pace, labs provide a better opportunity to raise your practical questions that are of general interest. Lab sessions are led primarily by TAs.

Web-site:

Notes and other materials are available on the Epid 718 (269) class web-site at <http://blackboard.unc.edu> and therefore there will generally be no handouts distributed in class (except for last minute additions and materials not easily or legally posted to the web-site). If you have difficulty printing these materials, notify a TA.

Lecture notes and readings are primarily distributed as Adobe Acrobat files (i.e., with .pdf extensions). Homework, project and exam assignments will be distributed as Word files so that you can use the files as a template for typing your assignments.

Required Texts:

Modern Epidemiology (2nd Edition)

Kenneth J. Rothman and Sander Greenland

Lippincott-Raven Publishers, 1998

ISBN: 0-316-75780-2

Statistical Analysis of Epidemiologic Data (3rd Edition)

Steve Selvin

Oxford University Press, 2004

ISBN: 0-195-17280-9

Homeworks:

There are two (2) homework assignments during the semester:

Homework #1

Assigned: **09/01/06** Due: **09/15/06** (in class)

Homework #2

Assigned: **10/18/06** Due: **11/10/06** (in class)

Homework assignments must be turned in during class on the day they are due.

Final Exam:

There is no mid-term exam for this course. The in-class final exam will be held from 12:00 noon to 3:00 PM on Monday December 11th, 2006.

Individual Consultations:

TAs will be available to conduct individual consultations on data analysis for projects and homework assignments. For help in understanding lecture material, you may seek an appointment with one of the TAs or with one of the faculty instructors.

The Project:

Each student is responsible for conducting a multifaceted analysis of a chosen research question, using methods and concepts taught in the course. Data can be from case-control, cohort or cross-sectional studies. The outcome variable of interest must be dichotomous, but you may select a continuous variable and categorize it into a binary variable based on some meaningful cut-point in order to model the probability of the event using an appropriate regression model.

The analyses should not be blind exercise, but should be motivated by some understanding of the topic area. If the topic is new to you, be sure to become familiar with the concepts (this does not mean you need to conduct a comprehensive literature review, but you should read at least a few relevant papers). Next, make certain you understand the design, objectives, and data collection protocol of the parent data set you will be using. After you understand the design and are familiar with the variables available, formulate a research question. Within the data set, identify a main “exposure” or predictor variable of interest (it can be continuous, categorical, or binary), a binary or survival time outcome (you may derive this by dichotomizing a continuous variable), at least six covariates that are potential confounders or effect modifiers, with the goal of using these variables to address the research question. Among the exposure and covariates, there should be at least one variable that is continuous and another that is either continuous or categorical with at least three levels.

The data set should include at least 300 observations. Extremely large data sets may be cumbersome, and so you may want to subset these for your own ease of use. However, more important than the total number of observations is the size of the cells. In particular, after you dichotomize your exposure and your outcome, the smallest cell in the resulting 2x2 table (usually the cell representing those both exposed and diseased) should be no less than 50 observations.

If you have not identified a data set by now, you should probably consider dropping the course!!!

Begin cleaning your data set immediately, and familiarizing yourself with coding and variable definitions. If there is some deficiency to your chosen data set (e.g. many missing values, undefined variables, etc.) it is best to find out about this sooner rather than later.

<u>LAB PROJECT</u>	<u>DUE</u>
Part 1	09/29/06 (in class)
Part 2	10/18/06 (in class)
Part 3	12/01/06 (in class)

Projects must be turned in during class on the day they are due.

Grading:

Some collaboration in the completion of homework assignments may be beneficial, but use good judgment in preventing your collaboration from becoming detrimental to your learning of the material. Submitted assignments should be your individual effort, even if you consult with other students about your strategy for obtaining these solutions.

All assignments must be submitted in printed (hard-copy) format.
No e-mail attachments will be accepted.

Late assignments carry a very severe penalty of 10% off per day late. Pleas of mercy for extenuating circumstances will be accepted only with written documentation. In previous years, a number of students suffered dramatically lower grades because of tardiness in submitting their assignments. Taking extra time to do a better job is probably not a worthwhile strategy, because the late penalty is so “expensive.”

Grading Complaints: If you have a complaint about a grade or grading of a question, submit the complaint in writing and I will consider it. Grading may potentially change in either direction as result of this reconsideration.

Homework:

Hw #1:	15%
Hw #2:	15%
<hr/>	
	30%

Final Exam:

	<hr/>
	25%

Project:

Part 1:	10%
Part 2:	15%
Part 3:	20%
<hr/>	
	45%

Total:	<hr/>
	100%

Final grades are based on the standard graduate school scale (H,P,L,F). The breakdown typically is something approximating the following scale, although allowances for observed distribution are made:

H: > 93%; P: 80 – 93%; L: 65 – 79%; F: < 65%.

No incompletes will be given except under special circumstances to be agreed upon by the student and instructor. Historically, about 25% of students have obtained a grade of H. A number of students have dropped the course in order to avoid a grade of L.

Syllabus:

Week 1	<u>23-Aug</u>	<u>25-Aug</u>	
11:00-11:50	INTRODUCTION	LEC1	
12:00-12:50		LAB1	
Week 2	<u>30-Aug</u>	<u>1-Sep</u>	HW 1 ASSIGNED FRIDAY 1 SEPT
11:00-11:50	LEC2	LEC3	
12:00-12:50		LAB2	
Week 3	<u>6-Sep</u>	<u>8-Sep</u>	
11:00-11:50	LEC4	LEC5	
12:00-12:50		LAB3	
Week 4	<u>13-Sep</u>	<u>15-Sep</u>	HW 1 DUE FRIDAY 15 SEPT
11:00-11:50	LEC6	LEC7	
12:00-12:50		LAB4	
Week 5	<u>20-Sep</u>	<u>22-Sep</u>	
11:00-11:50	LEC8	LEC9	
12:00-12:50		LAB5	
Week 6	<u>27-Sep</u>	<u>29-Sep</u>	PROJ PT 1 DUE FRIDAY 29 SEPT
11:00-11:50	LEC10	LEC11	
12:00-12:50		LAB6	
Week 7	<u>4-Oct</u>	<u>6-Oct</u>	
11:00-11:50	LEC12	LEC13	
12:00-12:50		LAB7	
Week 8	<u>11-Oct</u>	<u>13-Oct</u>	
11:00-11:50	LEC14	LEC15	
12:00-12:50		LAB8	
Week 9	<u>18-Oct</u>	<u>20-Oct</u>	PROJ PT 2 DUE WED 18 OCT
11:00-11:50	LEC16	FALL	HW 2 ASSIGNED WED 18 OCT
12:00-12:50		BREAK	
Week 10	<u>25-Oct</u>	<u>27-Oct</u>	
11:00-11:50	LEC17	LEC18	
12:00-12:50		LAB9	
Week 11	<u>1-Nov</u>	<u>3-Nov</u>	
11:00-11:50	LEC19	LEC20	
12:00-12:50		LAB10	
Week 12	<u>8-Nov*</u>	<u>10-Nov</u>	HW 2 DUE 10 NOV
11:00-11:50	LEC21	LEC22	
12:00-12:50		LAB11	
Week 13	<u>15-Nov</u>	<u>17-Nov</u>	
11:00-11:50	LEC23	LEC24	
12:00-12:50		LAB12	
Week 14	<u>22-Nov</u>	<u>24-Dec</u>	
11:00-11:50	THANKS	THANKS	
12:00-12:50	GIVING	GIVING	
Week 15	<u>29-Nov</u>	<u>1-Dec</u>	PROJ PT3 DUE 1 DEC
11:00-11:50	LEC25	LEC26	
12:00-12:50		LAB13	
Week 16	<u>6-Dec</u>		
11:00-11:50	LEC27		
12:00-12:50			

* APHA Annual Meeting: Nov 4-8

Epidemiology 718 (269) Fall 2006

LECTURES	TOPIC	
1	25-Aug	Review of Logs, Logits, OR and Linear Models DR
2	30-Aug	Exploratory Data Analysis DR
3	1-Sep	A General Approach to Linear Models DR
4	6-Sep	Logistic Regression in Cohort Context: Introduction JK
5	8-Sep	Multiple Logistic Regression in Cohort Context JK
6	13-Sep	Coefficient Interpretation in Logistic Regression JK
7	15-Sep	Dose Response Trend in Logistic Regression JK
8	20-Sep	Maximum Likelihood Theory JK
9	22-Sep	Comparing Models and Constructing Intervals JK
10	27-Sep	Model Building JK
11	29-Sep	Effect and Interaction Analysis JK
12	4-Oct	Binomial (Relative Risk) Regression Miller
13	6-Oct	Bayesian Logistic Regression MacLehose
14	11-Oct	Logistic Regression in Case-Control Context: Intro DR
15	13-Oct	Matched Case-Control-Data DR
16	18-Oct	Analysis of Matched Case-Control Data DR
17	25-Oct	Defining Time-Related Factors DR
18	27-Oct	Statistical Interactions involving matching factors DR
19	1-Nov	Pooling Strata DR
20	3-Nov	Nested case-control studies DR
21	8-Nov	Worked examples DR
22	10-Nov	Ordinal and Multinomial Logistic Regression DR
23	15-Nov	Assessing Model Fit JK
24	17-Nov	Clustered Binary Data JK
25	29-Nov	Cluster Specific vs Population Average Models JK
26	1-Dec	Nonparametric methods (smoothers) JK
27	6-Dec	Summary and Retrospective JK

LABS:		
1	25-Aug	Management of Data and Univariate Analysis (LV/SN)
2	1-Sep	Graphical Methods and Exploratory Data Analysis (TAs)
3	8-Sep	Confounding and Joint Effects (TAs)
4	15-Sep	Dose/Response (TAs)
5	22-Sep	Model Building (TAs)
6	29-Sep	Review of Regression Slopes and P-Values (JK)
7	6-Oct	Effect and Interaction Analysis (TAs)
8	13-Oct	Conditional Logistic Regression and Matched Analysis (DR)
9	27-Oct	Review Sample Projects From Previous Years (TAs)
10	3-Nov	Data Preparation & Analysis of Nested Case-Control Studies (DR)
11	10-Nov	Ordinal and Multinomial Regression (DR/JK)
12	17-Nov	Regression Diagnostics & Clustered Binary Data Models (SM)
13	1-Dec	Nonparametrics, smoothers (JK)