Statistical Methods for Correlated Data

PubH 7430
3 credits, A/F or S/N

Fall 2005 – SYLLABUS

Instructor
Lynn E. Eberly, Associate Professor
E-mail lynn@biostat.umn.edu
Phone 612-624-1436
Office Mayo A-465
Office hours TBA
Campus Mail MMC 303

Teaching Assistant
Luping Zhao, PhD Student
E-mail lupingz@biostat.umn.edu
Phone 612-624-0131
Office Mayo A-452
Office hours TBA

Course Objectives
Correlated data often arise from data collected over time or space, group randomizations, cluster sampling, nested designs, or random effects assumptions. This course will cover two classes of models appropriate for these settings: general linear (mixed) models for normally distributed data and generalized linear (mixed) models for non-normally distributed data, such as binomial, Poisson, and gamma. We will discuss graphical data exploration, correlation structures, parameter estimation / testing / inference, model selection and interpretation, diagnostics, and model limitations. Modeling will be done using SAS.

Class Meetings
Class meetings will be a mixture of lecture and discussion of applications. The instructor's lecture notes will be available for download from the course web page (see Miscellaneous Notes below). We will meet Tuesdays and Thursdays 11:15 am – 12:30 pm in Mayo D-327.

Prerequisites
Statistics at the level of PubH 6451 or PubH 7406 or Stat 5303; and SAS programming at the level of PubH 6420; and familiarity with matrix notation; or permission of instructor.

Guidelines
Students should be familiar with the basic notions of random variables, statistical inference, multiple linear regression, and logistic regression. Familiarity with matrix notation is helpful; we will review this VERY briefly at the beginning of the course and make use of matrix notation throughout. The course is meant to be accessible to graduate students in all fields. The underlying statistical theory will be outlined but not stressed and the main focus will be on applications.
### Required Materials


### Recommended Materials


### Computing

All statistical modeling will be done in SAS. All students can request access to SAS on the Biostatistics computer UNIX network. Students with access to SAS elsewhere are free to use whichever system is most convenient. SAS Version 9 will be needed. Graphing of data can be done in the student’s package of choice.

### Work Expectations

Students are expected to attend class, participate in class discussions, and complete the assigned homework, exam, and project. Students are also expected to complete assigned reading from the required materials.

### Homework

There will be five homework assignments during the semester. We encourage you to work together in computing and discussing the problems. However, each student is expected to independently write up the submitted assignment using her or his own computing and giving explanations in her or his own words. All assignments will involve computing; please attach only relevant computer output to what you turn in. Some assignments may also include reading and writing about a related journal article.

You will get about two weeks to work on each homework assignment. The homework will be graded on a scale of 0 to 25 points, for a maximum of 125 points possible over the semester. Late homework will lose three points per day, unless arrangements have been made with the instructor prior to the due date.

### Exam

There will be one in class open-book / open-notes exam towards the middle of the semester. The exam is worth 100 points.

### Project

There will be a final project worth 100 points towards the end of the semester (no final exam). The project could take one of two structures:

1. Obtain a data set with a correlated structure; using relevant background information, determine several scientific questions to be answered by the data. Carry out a full analysis that addresses these scientific questions using any appropriate modeling strategies covered in class and discuss the results. **An oral report is required.**

2. Choose an area of correlated data analysis that is not covered in this course. Review and summarize the relevant literature. Or, carry out a simulation to demonstrate or explore a concept. **An oral report is required.**
More details will be handed out in mid October. Each student must write a <1/2 page project proposal (by e-mail is fine) and get approval from the instructor by early November. We will have a few data sets available for option (1) and a few ideas available for option (2) if you do not already have something you would like to work on. No written report is required.

**Grading**

A letter grade will be determined from the percentage of (325 possible) points each student receives as follows:

- B+ 87-89%
- C+ 77-79%
- D+ 67-69%
- A 93-100%
- B 83-86%
- C 73-76%
- D 63-66%
- A- 90-92%
- B- 80-82%
- C- 70-72%
- F 0-62%

For those enrolled S/N, a letter grade of C or better must be achieved to receive an S. The University Senate has established a uniform grading policy for all letter grades: [www1.umn.edu/usenate/policies/gradingpolicy.html](http://www1.umn.edu/usenate/policies/gradingpolicy.html). If you would like to switch grading options (e.g., A/F to S/N), it must be done within the first two weeks of the semester.

**Outline**

All sections include applications. Numbers in () are the week numbers.

- Introduction and some examples (week 1)
- Graphical exploration of correlated data (1-2)
- Matrix algebra and review of regression in matrix notation (2)
- Classical methods for normally distributed repeated measures (3)
  - derived variables models
  - repeated measures analysis of variance
  - drawbacks and limitations of classical methods
- Methods for normally distributed correlated data (4-11)
  - general linear (mixed) models
  - estimation / testing / interpretation / model building / diagnostics
  - “time-varying” covariates / parameterization issues
- Review of methods for non-normally distributed independent data (11)
  - logistic regression
  - Poisson regression
- Methods for non-normally distributed correlated data (12-13)
  - generalized linear mixed models
  - estimation / testing / interpretation / model building / diagnostics
  - generalized estimating equations (GEE)
  - estimation / testing / interpretation / model building / diagnostics
- Advanced topics (time permitting) (14)
  - missing data issues
  - computational issues

**Resources**

Several books on background material and further references on the material we will cover are on reserve in the Bio-Medical Library Reserve Desk (Diehl Hall 270):

1. Diggle, Heagerty, Liang, and Zeger (2002). *Analysis of Longitudinal Data*, London: Oxford University Press. For much more theoretical detail on all of the material we will cover.
(2) Davis (2002). *Statistical Methods for the Analysis of Repeated Measurements*, New York: Springer-Verlag. For a bit more theoretical detail and similar application detail on all of the material we will cover.


Homework solutions and previous years’ exams and projects will be on reserve as well. In addition, the Biostatistics Reading Room (Mayo A-460) has full documentation for SAS Version 8, two books on graphing in SAS, and the following:


These cannot be checked out except to make copies, but can be browsed in the Reading Room.

**Miscellaneous Notes**

Clarifications from class, changes in homework assignments, or course announcements will be distributed by e-mail to your University x500 account. Also, we will be maintaining a course web page at

http://www.biostat.umn.edu/~lynn/ph7430.html

Lecture notes, some computing information and code, data sets, and office hours will be available there. At the end of the semester, please take seriously the opportunity to give feedback in the course evaluations and advise us on how the course might be improved.

The following policies are common to all courses in the School of Public Health and are mandated by the School’s Educational Policies document.

**Grading Option Policy**

If applicable, students may change grading options during the initial registration period or during the first two weeks of the term. **The grading option may not be changed after the second week of the term.**

An incomplete grade is permitted only in cases of extraordinary circumstances and following consultation with the instructor. In such cases, an “I” grade will require a specific written agreement between the instructor and the student specifying the time and manner in which the student will complete the course requirements. Extension for completion of the work will not exceed one year.

**Scholastic Dishonesty**

Students are responsible for knowing the University of Minnesota Board of Regents’ policy on student conduct and scholastic dishonesty: www1.umn.edu/regents/policies/academic/StudentConductCode.pdf. Scholastic dishonesty as defined in the policy will be reported
to the Office of Student Judicial Affairs: www.sja.umn.edu and will result in a grade of “F” or “N” for the entire course.

Plagiarism is an important element of this policy. It is defined as the presentation of another’s writing or ideas as your own. Serious, intentional plagiarism will result in a grade of “F” or “N” for the entire course. For more information on this policy and for a helpful discussion of preventing plagiarism, please consult University policies and procedures regarding academic integrity: writing.umn.edu/tww/plagiarism/.

Students are urged to be careful that they properly attribute and cite others’ work in their own writing. For guidelines for correctly citing sources, go to tutorial.lib.umn.edu and click on “Citing Sources.”

In addition, original work is required in this course. It is unacceptable to hand in assignments for this course for which you receive credit in another course unless by prior agreement with the instructor. Building on a line of work begun in another course or leading to a thesis, dissertation, or final project is acceptable.

If you have any questions, consult the instructor.

Withdrawal Policy

School of Public Health students may withdraw from a course through the second week of the semester without permission. No “W” will appear on the transcript. After the second week, students are required to do the following:

1) The student must contact and notify their advisor and course instructor informing them of the decision to withdraw from the course.

2) The student must send an e-mail to the SPH Student Services Center (SSC) at sph-ssc@umn.edu. The email must provide the student name, ID#, course number, section number, semester, and year with instructions to withdraw the student from the course, and acknowledgement that the instructor and advisor have been contacted.

3) The advisor and instructor must e-mail the SSC acknowledging the student is canceling the course. All parties must be notified of the student’s intent.

4) The SSC will complete the process by withdrawing the student from the course after receiving all e-mails (student, advisor, and instructor). A “W” will be placed and remain on the student transcript for the course.

5) After discussion with their advisor and notification to the instructor, students may withdraw up until the eighth week of the semester. There is no appeal process.

It is University policy to provide, on a flexible and individualized basis, reasonable accommodations to students who have a documented disability (e.g., physical, learning, psychiatric, vision, hearing, or systemic) that may affect their ability to participate in course activities or to meet course requirements. Students with disabilities are encouraged to contact Disability Services to have a confidential discussion of their individual needs for accommodation. Disability Services is located in Suite 180 McNamara Alumni Center, 200 Oak Street. Staff can be reached by calling (612) 626-1333 (V/TTY).