SYLLABUS & COURSE INFORMATION

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

PUBH 7430, SECTION 001

Statistical Methods for Correlated Data
Fall 2019

COURSE & CONTACT INFORMATION

Credits: 3
Meeting Day(s): Tuesday and Thursday
Meeting Time: 9:45 - 11:00 AM
Meeting Place: Moos Health Sciences Tower 2-520

Instructor: Erika Helgeson, PhD
Email: helge@umn.edu
Office Phone: 612-626-8156
Office Hours: Mayo A454-1:Tuesdays from 11:15 AM - 12:15 PM, or by appointment
Office Location: For office hours: Mayo A454-1; Otherwise: 2221 University Ave. SE, Room 361

TAs: To be announced
Office Hours: To be announced

COURSE DESCRIPTION

Correlated data arise in many situations, particularly when observations are made over time and space or on individuals who share certain underlying characteristics. In this course, we will study techniques for exploring and describing correlated data, along with statistical methods for estimating population parameters (mostly means) from these data. We will focus primarily on the class of generalized linear models (both with and without random effects) for normally and non-normally distributed data. Wherever possible, techniques will be illustrated using real-world examples. Computing will be done using R and SAS.

PUBH 7430 is a course designed for second year MS students in Biostatistics or Statistics and PhD students in other quantitative fields. There will be a focus on applications, although underlying statistical concepts and theory will be reviewed as necessary.

COURSE PREREQUISITES

Statistics: Regression at the level of PubH 6451 (Biostatistics II) or PubH 7405 (Biostatistics Regression) or Stat 5302 (Applied Regression Analysis). Students should be familiar with the basic notions of random variables, statistical inference (confidence intervals, hypothesis testing), multiple linear regression, and logistic regression.

Linear Algebra: Some familiarity with basic matrix notation and operations (multiplication, inverse, transpose). We will review this VERY briefly towards the beginning of the course. During the semester, the underlying statistical theory will be outlined using matrix notation, but deep understanding of the theory is not necessary for homework or exams.

Computing: Working knowledge of SAS or R at the level of PubH 6420. Other software (e.g., Stata or SPSS) can be used for data analysis but support for these packages cannot be guaranteed from the instructor or teaching assistant.

COURSE GOALS & OBJECTIVES

Upon completing this course, students should be able to:

- Identify situations where correlated data may arise.
- Describe and summarize correlation in a dataset both graphically and numerically.
- Apply appropriate statistical estimation techniques to answer scientific questions using correlated data.
- Understand both the strengths and limitations of these techniques.
- Communicate written analysis results in a manner appropriate for a scientific publication.
- Clearly present a scientific problem, analysis, and results using a poster presentation.
- Work as a team to answer a scientific question.

METHODS OF INSTRUCTION AND WORK EXPECTATIONS

Students are expected to attend class, participate in class discussions, and complete the assigned homework, exam, and project. Working together on homework assignments is permitted, but **each student is expected to independently write up homework assignments, including any code, in their own words.**

Email
I am happy to answer email questions when time permits, but due to other commitments, I cannot always reply immediately. Please post questions of general interest to the Discussion Board on Canvas. Email questions that are not of general interest should be sent to me **and** both TAs. Email is not often the best option for getting questions answered: it is usually more efficient to attend my or the TAs' office hours.

Course Workload Expectations
PubH 7430 is a 3-credit course. The University expects that *for each credit*, you will spend a minimum of three hours per week attending class or comparable online activity, reading, studying, completing assignments, etc. over the course of a 15-week term. Thus, this course requires approximately 135 hours of effort spread over the course of the term in order to earn an average grade.

Learning Community
This course will ask students to work in teams or discussion groups. We do not come to our courses with identical backgrounds and experiences and building on what we already know about collaborating, listening, and engaging is critical to successful professional, academic, and scientific engagement with topics.

In this course, students are expected to engage with each other in respectful and thoughtful ways.

In group work, this can mean:

- Setting expectations with your groups about communication and response time as soon as groups are assigned and contacting the TA or instructor if scheduling problems cannot be overcome.
- Setting clear deadlines and holding yourself and each other accountable.
- Determining the roles group members need to fulfill to successfully complete the project on time.
- Developing a rapport prior to beginning the project (what prior experience are you bringing to the project, what are your strengths as they apply to the project, what do you like to work on?)

In group discussion, this can mean:

- Respecting the identities and experiences of your classmates.
- Avoid broad statements and generalizations. Group discussions are another form of academic communication and responses to instructor questions in a group discussion are evaluated. Apply the same rigor to crafting discussion posts as you would for a paper.
- Consider your tone and language, especially when communicating in text format, as the lack of other cues can lead to misinterpretation.

Like other work in the course, all student to student communication is covered by the Student Conduct Code (https://z.umn.edu/studentconduct).

COURSE TEXT & READINGS

There is no required text for this course. However, lecture notes will draw heavily from:

Fitzmaurice, Laird, and Ware (2004) Applied Longitudinal Analysis. John Wiley and Sons.
This text provides an introductory presentation of longitudinal data methods suitable for graduate level work with applications in SAS.

And:

Diggle, Heagerty, Liang, and Zeger (2002) Analysis of Longitudinal Data, New York: Oxford University Press. ISBN 0198524846.
This text gives some mathematical theory as well as practical aspects and applications of methods for the analysis of longitudinal data. Many of the technical details it provides are beyond the scope of this course, but you may find it to be a useful reference.

The following books may also be of use:

Verbeke and Molenberghs (2000). *Linear Mixed Models for Longitudinal Data*, New York: Springer-Verlag, Inc. ISBN 0387950273.
If you expect your future work to involve lots of longitudinal data analysis in SAS, this is a helpful book to own.

The Biostatistics Reading Room (Mayo A-460) has full documentation for SAS, two books on graphing in SAS, and the following:

Littell, Miliken, Stroup, and Wolfinger (1996). *SAS System for Mixed Models*, Cary, NC: SAS Institute, Inc. For detailed examples using SAS `PROC MIXED`.

Khattree and Naik (1999). *Applied Multivariate Statistics with SAS Software*, 2nd Edition, Cary, NC: SAS Institute, Inc. For detailed graphing and modeling examples using SAS.

COURSE OUTLINE/WEEKLY SCHEDULE

Tentative schedule of topics:

- Overview and introduction to correlated data (**Week 1**: Sept 3, 5)
- Exploratory and descriptive analyses (**Week 2**: Sept 10, 12)
- Generalized linear models (GLMs) for independent data (**Week 3**: Sept 17, 19; **Week 4**: Sept 24, 26; **Week 5**: Oct 1)
- Marginal models and generalized estimating equations (**Week 5**: Oct 3; **Week 6**: Oct 8, 10; **Week 7**: Oct 15, 17)
- Mixed effects models (**Week 8**: Oct 22, Oct 24; **Week 9**: Oct 29, 31; **Week 10**: Nov 5, Nov 7; **Week 11**; Nov 12, 14)
- Special topics (time permitting: **Week 12:** Nov 21; **Week 13**: Nov 26; **Week 14**: Dec 3, 5)

Tentative schedule of activities:

Note that there will not be an in-class final exam for this course. Instead, there will be a group project that will be worked on throughout the semester; a final report for the project is due on Wednesday, December 11 (the day following the last day of class). Additionally, students will be individually peer reviewing another group's final report. This peer review will need to be electronically submitted by the end of the scheduled final exam time for the course (10 AM on Thursday, December 19). Thus, the last in-person activity for the course will be the last day of the course (Tuesday, December 10).

Activity/Assignment	**Due Date**	**Due Time**
Project: Survey	Fri. Sept. 6	6 PM
Assignment 0	Tues. Sept. 10	In Class (No grade)
Assignment 1	Thurs. Sept. 26	6 PM
Project: Project proposal	Thurs., Oct. 3	6 PM
Assignment 2	Thurs., Oct. 10	6 PM
Assignment 3	Thurs., Oct. 24	6 PM
Project: Statistical analysis plan	Thurs., Nov. 7	6 PM
Assignment 4	Thurs., Nov. 14	6 PM
Midterm	**Tues., Nov. 19**	In Class
Project: Results	Mon Dec. 2	6 PM
Assignment 5	Thurs., Dec. 5	6 PM
Project: Poster presentations	Tues., Dec. 10	In Class
Project: Final report	Wed., Dec. 11	6 PM
Project: Peer evaluation	Thurs., Dec. 19	10 AM

SPH AND UNIVERSITY POLICIES & RESOURCES

The School of Public Health maintains up-to-date information about resources available to students, as well as formal course policies, on our website at www.sph.umn.edu/student-policies/. Students are expected to read and understand all policy information available at this link and are encouraged to make use of the resources available.

The University of Minnesota has official policies, including but not limited to the following:

- Grade definitions
- Scholastic dishonesty
- Makeup work for legitimate absences
- Student conduct code
- Sexual harassment, sexual assault, stalking and relationship violence
- Equity, diversity, equal employment opportunity, and affirmative action
- Disability services
- Academic freedom and responsibility

Resources available for students include:

- Confidential mental health services
- Disability accommodations
- Housing and financial instability resources
- Technology help
- Academic support

EVALUATION & GRADING

There will be five homework assignments, distributed roughly every two weeks. Homework keys will be posted on Canvas and the TAs will return graded assignments within one week of the assignment deadline.

There will be an in-class, closed-book midterm exam at approximately two-thirds of the way through the semester.

Students will also complete a group project. The project grade will be based on several components completed throughout the semester including: project proposal (10 points), statistical analysis plan (20 points), results (20 points), poster presentation (20 points), final written report (40 points), reproducibility of analysis (20 points), peer review of final project (10 points), and active and timely participation in the project (15%), as determined by group member feedback and by the instructor's judgement based on such things as your participation in the discussion of the project presentations. The individual components will be graded by the instructor and returned with feedback within one week of the project deadline.

The final course grade will be determined by equally weighting the homework, midterm, and group project grades:

- 33% homework (each assignment contributes equally)
- 33% midterm
- 33% group project

No make-up work will be allowed for missed homework assignments, project assignments, or the midterm. Exceptions will be made for legitimate absences (as defined in the University's policy) and arrangements for these should be made with the instructor as soon as possible. Please refer to the University's policy (https://policy.umn.edu/education/makeupwork) for further details.

Grading Scale
A letter grade will be determined from the percentage of points each student receives. **The following translation of number grades to letters represents the strictest possible scheme that could be used**; the instructor may make the scheme less strict (e.g., A = 91-100%, A- = 88-90%, etc.) depending on the final distribution of numerical grades:

% In Class	Grade	GPA
93 - 100%	A	4.000
90 - 92%	A-	3.667
87 - 89%	B+	3.333
83 - 86%	B	3.000
80 - 82%	B-	2.667
77 - 79%	C+	2.333
73 - 76%	C	2.000
70 - 72%	C-	1.667
67 - 69%	D+	1.333
63 - 66%	D	1.000
< 63%	F	

- A = achievement that is outstanding relative to the level necessary to meet course requirements.
- B = achievement that is significantly above the level necessary to meet course requirements.
- C = achievement that meets the course requirements in every respect.
- D = achievement that is worthy of credit even though it fails to meet fully the course requirements.
- F = failure because work was either (1) completed but at a level of achievement that is not worthy of credit or (2) was not completed and there was no agreement between the instructor and the student that the student would be awarded an I (Incomplete).
- S = achievement that is satisfactory, which is equivalent to a C- or better
- N = achievement that is not satisfactory and signifies that the work was either 1) completed but at a level that is not worthy of credit, or 2) not completed and there was no agreement between the instructor and student that the student would receive an I (Incomplete).

Evaluation/Grading Policy	Evaluation/Grading Policy Description
Scholastic Dishonesty, Plagiarism, Cheating, etc.	You are expected to do your own academic work and cite sources as necessary. Failing to do so is scholastic dishonesty. Scholastic dishonesty means plagiarizing; cheating on assignments or examinations; engaging in unauthorized collaboration on academic work; taking, acquiring, or using test materials without faculty permission; submitting false or incomplete records of academic achievement; acting alone or in cooperation with another to falsify records or to obtain dishonestly grades, honors, awards, or professional endorsement; altering, forging, or misusing a University academic record; or fabricating or falsifying data, research procedures, or data analysis (as defined in the Student Conduct Code). For additional information, please see https://z.umn.edu/dishonesty The Office for Student Conduct and Academic Integrity has compiled a useful list of frequently asked questions pertaining to scholastic dishonesty: https://z.umn.edu/integrity. If you have additional questions, please clarify with your instructor. Your instructor can respond to your specific questions regarding what would constitute scholastic dishonesty in the context of a particular class-e.g., whether collaboration on assignments is permitted, requirements and methods for citing sources, if electronic aids are permitted or prohibited during an exam. Indiana University offers a clear description of plagiarism and an online quiz to check your understanding (http://z.umn.edu/iuplagiarism).
Late Assignments	No make-up work will be allowed for missed homework assignments, project assignments, or the midterm. Exceptions will be made for legitimate absences (as defined in the University's policy) and arrangements for these should be made with the instructor as soon as possible. Please refer to the University's policy (https://policy.umn.edu/education/makeupwork) for further details.
Attendance Requirements	Students are expected to attend class, participate in class discussions, and complete the assigned homework, exam, and project. If students must miss a class, they are expected to make arrangements with a classmate to find out what was missed in order to not fall behind in the course.
Extra Credit	Check-in quizzes on Canvas will be available after some class periods (due at 11:59 pm on the day of that class) to help the instructor understand which topics students are finding most challenging. These quizzes are optional but completing all of them will result in a one percentage point increase in the student's final course grade. Completing a portion of these quizzes will result in a proportionate percentage point increase.
Regrading	Student requests for regrading of assignments and exams must be made to the instructor **in writing within three days** of when the graded assignment or exam is returned to students. Regrading will only be considered for issues involving **two points or more**. Please email the instructor your request, including the question number, which points you feel were unfairly deducted, and the reason. The instructor will exclusively handle all regrading, so please do not ask the TAs to do this. If a student's grade is changed, all students will be notified and can request additional points if they had similar answers that were penalized.

CEPH COMPETENCIES

Competency	Learning Objectives	Assessment Strategies
Select quantitative and qualitative data collection methods appropriate for a given health context	• Identify situations where correlated data may arise. • Apply appropriate statistical estimation techniques to answer scientific questions using correlated data. • Understand both the strengths and limitations of these techniques.	• Homework assignments • Mid-term exam • Group project
Analyze quantitative and qualitative data using biostatics, informatics, computer-based programming and software as appropriate	• Describe and summarize correlation in a dataset both graphically and numerically. • Apply appropriate statistical estimation techniques to answer scientific questions using correlated data.	• Homework assignments • Group project
Interpret results of data analysis for public health research, policy of practice	• Communicate analysis results in a manner appropriate for a scientific publication.	• Homework assignments • Mid-term exam • Group project
Communicate audience-appropriate public health content, both in writing and through oral presentation	• Communicate written analysis results in a manner appropriate for a scientific publication. • Clearly present a scientific problem, analysis, and results using a poster presentation.	• Group project
Perform effectively on interprofessional teams	• Work as a team to answer a scientific question.	• Group project