SYLLABUS & COURSE INFORMATION

SCHOOL OF PUBLIC HEALTH
UNIVERSITY OF MINNESOTA

PUBH 7402, SECTION 001

Biostatistics Modeling and Methods
Spring 2019

COURSE & CONTACT INFORMATION

Credits: 4
Meeting Day(s): Tu & Th
Meeting Time: 12:20 - 2:15 pm
Meeting Place: Moos 2-690

Instructor: Dr. Lynn Eberly & Dr. Wei Pan
Email: leberly@umn.edu (Lynn) & panxx014@umn.edu (Wei)
Office Phone: 612-624-1436 (Lynn) & 612-626-2705 (Wei)
Fax: 612-626-0660
Office Hours: See the Canvas course site
Office Location: See the Canvas course site

COURSE DESCRIPTION

This course is the second part of a two-course sequence intended for PhD students in the School of Public Health who need a rigorous approach to probability and statistics and statistical inference with applications to research in public health. This course is about modeling data and using the estimates of those models along with estimates of uncertainty of those models to answer scientific research questions. Students are allowed to choose whichever software they prefer.

Acknowledgments
The contents of PubH 7402 have been developed with the contributions of numerous instructors. The notes follow those of Dr. Sandra Safo and Dr. Weihua Guan, with some modifications. Former faculty Dr. Melanie Wall had roles in both the conceptual development and actual content of the current course, and is acknowledged for her contributions.

COURSE PREREQUISITES

PubH 7401 is required; or permission by instructors.

COURSE GOALS & OBJECTIVES

Upon completion of this course, the student should understand the fundamentals of generalized linear models for continuous, ordinal and categorical data. The students should also understand how to model data coming from correlated sampling structures and how to deal with time-to-event data. This means comprehending the concepts of model construction, applying these concepts to actual datasets, implementing the appropriate model in standard statistical software, and understanding the interpretation of those models in the context of those datasets.

METHODS OF INSTRUCTION AND WORK EXPECTATIONS

Course Workload Expectations
PubH 7402 is a 4 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 4*45=180 hours of effort spread over the course of the term in order to earn an average grade.

Instruction will be in-class presentation of concepts with discussion of definitions, implementation, and interpretation. Examples and applications will come from the health science related fields. 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, using their own computing and in their own words.**

© 2018 Regents of the University of Minnesota. All rights reserved. The University of Minnesota is an equal opportunity educator and employer. Printed on recycled and recyclable paper with at least 10 percent postconsumer waste material. This publication/material is available in alternative formats upon request to 612-624-6669.

Course Software
The course will present all methods using SAS. STATA tips will be included in some situations and is the software used in the primary text book by Vittinghoff et al. (shown below). R/RStudio also provides comprehensive modeling functionality, and instructors and TA are well versed in R; R will also be presented in some situations. R and SAS is the software used in the secondary text book by Le and Eberly (shown below). All of these software tools are comprehensive statistical tools, capable of performing most if not all of the traditional and the recently developed statistical methodology. Students are welcome to choose any other software to perform analyses for homework in this course, and can change their software use at any time.

You can find information on software available at the University of Minnesota academic computing labs here: https://it.umn.edu/service-details/academic-computer-labs. You can also obtain a license for SAS or Stata through the University of Minnesota https://it.umn.edu/students. R is free software and can be downloaded here: https://www.r-project.org/.

Useful tutorials at the UCLA website:
SAS: https://stats.idre.ucla.edu/sas/
STATA: https://stats.idre.ucla.edu/stata/
R: https://stats.idre.ucla.edu/r/

Learning Community
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 the classroom, outside of the classroom, and in the context of the group project.

In group work, this can mean:

- Setting expectations with your groups about communication and response time during the first week of the semester (or 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?)

During in-class or group project settings, this can mean:

- Respecting the identities and experiences of your classmates.
- Avoid broad statements and generalizations. Group discussions and responses to instructor questions are another form of academic communication. Apply the same rigor to crafting discussion ideas 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-with-student communication is covered by the Student Conduct Code (https://z.umn.edu/studentconduct) and faculty-with-student communication is covered by the Faculty and Staff Conduct Code (https://regents.umn.edu/sites/regents.umn.edu/files/policies/Code_of_Conduct.pdf).

COURSE TEXT & READINGS

Primary text: Regression Methods in Biostatistics: Linear Logistic, Survival, and Repeated Measures Models. Vittinghoff, E., Glidden, D.V., Shiboski, S.C., and McCulloch, Second Edition. Springer. 2012. DOI: 10.1007/978-1-4614-1353-0. Hardcover ISBN: 978-1-4614-1352-3.

(EBook: http://link.springer.com.ezp1.lib.umn.edu/book/10.1007%2F978-1-4614-1353-0) Course reading in the weekly schedule below covers Chapters 1-8 and 10. This book presents Stata examples.

Secondary text: Introductory Biostatistics. Le C and Eberly LE, Second Edition. Wiley. 2016. Hardcover ISBN: 978-0470905401.

(EBook: http://login.ezproxy.lib.umn.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=nlebk&AN=1221491&site=ehost-live) Course reading in the weekly schedule below covers **Chapter 12 only**, but other chapters cover similar material to our primary text; this book presents SAS and R code examples.

COURSE OUTLINE/WEEKLY SCHEDULE **This schedule is approximate!** VGSM = Vittinghoff, Glidden, Shiboski, McCulloch. LE = Le, Eberly.

Week	Topic	Textbook Reading	Activities and Assignments Due
Week 1: 1/22, 1/24	• Course overview, General Linear Model	• VGSM Chapters 1-4	• Reading 1 due 1/24 in class
Week 2: 1/29, 1/31	• General Linear Model	• VGSM Chapter 4	• Nothing due (work on Final Project Step 1)
Week 3: 2/5, 2/7	• General Linear Model	• VGSM Chapter 4	• Assignment 1 due 2/8
Week 4: 2/12, 2/14	• Generalized Linear Model	• VGSM Chapter 5	• Reading 2 due 2/14 in class
Week 5: 2/19, 2/21	• Generalized Linear Model	• VGSM Chapters 5, 8	• Final Project Step 1 due 2/22
Week 6: 2/26, 2/28	• Generalized Linear Model	• VGSM Chapters 8	• Assignment 2 due 3/1
Week 7: 3/5, 3/7	• Generalized Linear Model, Model selection	• VGSM Chapters 8, 10	• Nothing due (work on Final Project Step 2)
Week 8: 3/12, 3/14	• Midterm exam; Longitudinal and clustered data	• VGSM Chapters 7.1-7.3	• Midterm 3/12 in class • Assignment 3 due 3/15
SPRING BREAK: 3/18 – 3/22	• No topic!	• No reading!	• No assignment!
Week 9: 3/26, 3/28	• Linear Mixed Effects Model	• VGSM Chapters 7.5, 7.6 • LE Chapter 12.1-12.2; slides	• Reading 3 due 3/28 in class
Week 10: 4/2, 4/4	• Generalized Linear Mixed Effects Model	• VGSM Chapter 7.7.4 • LE Chapter 12.3; slides	• Final Project Step 2 due 4/5
Week 11: 4/9, 4/11	• Generalized Estimating Equations (GEE)	• VGSM Chapter 7.4 • LE Chapter 12.4; slides	• Nothing due (work on Final Project Step 3)
Week 12: 4/16, 4/18	• GEE vs. mixed effects models; Survival Analysis: concepts	• VGSM Chapters 7.8; 3.5.1, 6.1 • LE Chapter 13.1; slides	• Assignment 4 due 4/19
Week 13: 4/23, 4/25	• Survival Analysis: one - and two-sample problems	• VGSM Chapters 3.5.2-3.5.6 • LE Chapter 13.2; slides	• Reading 4 due 4/25 in class • Final Project Step 3 due 4/26
Week 14: 4/30, 5/2	• Survival Analysis: Cox regression	• VGSM Chapters 6.2, (6.3—6.4) • LE Chapter13.3-13.4; slides	• Assignment 5 due 5/3
Week 15: 5/6 – 5/10	• NO CLASS	• --	• Nothing due (work on Final Project Steps 4, 5, and 6)
Week 16: 5/13 – 5/15	• NO CLASS. Official final exam time on 5/15, 10:30am-12:30pm, will be used for poster presentations.	• --	• Final Project Step 4 due 5/15 midnight • Final Project Step 5 due 5/15 10:20am • Final Project Step 6 due 5/15 midnight

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

Homework
Homework will be assigned approximately every 3 weeks, with a total of 5 assignments. Each assignment will be graded on a scale of 0 to 25 points (125 possible points over all assignments).

Exam
One midterm exam will be given in class and graded on a scale of 0 to 100 points. The midterm will cover material through Generalized Linear Models. There is no final exam; we will use the University-assigned official final exam time for poster presentations of the Final Project.

Final Project
The final project will be a group project where your group analyzes data of the group's choice. Your group will prepare a project including: (Step 1) a description of a proposed data set and scientific questions to be answered (15 points), (Step 2) a detailed statistical analysis plan (15 points), (Step 3) a summary of the analysis results (15 points), (Step 4) a polished short (5-10 page double-spaced) report of the group's work (25 points), and (Step 5) a poster presentation (20 points) of the group's work. As evidence of the reproducibility of the work, the group will also submit (Step 6) their data and well-documented analysis code (10 points). The final project in total is therefore graded on a scale of 0 to 100 points.

This will be a **group** project; students **must** form into groups of **2 or 3**. Group members should find a data set, and discuss and assign responsibilities, early in the semester; some guiding materials and resources for this allocation of responsibilities will be provided by the instructors. All members of a group will receive the same grade for all parts of the project. Once your group has formed and identified data of interest, you will need to meet briefly with the instructors to discuss your data. More final project information is provided in the Canvas web site.

Final Grade
The final grade will be determined as the percentage of points earned out of the total possible points (125+100+100=325). A letter grade will be determined from the percentage of points earned.

Grading Scale
The University uses plus and minus grading on a 4.000 cumulative grade point scale in accordance with the following table.

% 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
< 62%	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	Each homework is worth 25 points; 5 points will be lost per day late unless an extension has been arranged with an instructor. The final project is in total worth 100 points; 5 points will be lost per day late unless an extension has been arranged with an instructor. If you think you may need to turn something in late, please communicate with us promptly!!
Attendance Requirements	Students are expected to attend class. If you need to miss a class, please discuss this with the instructor, in advance whenever possible.
Extra Credit	None.

CEPH COMPETENCIES

Competency	Learning Objectives	Assessment Strategies
1. Apply epidemiological methods to the breadth of settings and situations in public health practice	When given a specific research context, recognize which statistical analysis method is appropriate (e.g., logistic regression analysis) to address the scientific or public health question for the given data set. Carry out diagnostics to assess whether the assumptions of the method hold for the given data set.	Homework assignments and the group project will cover this learning objective.
2. Select quantitative and qualitative data collection methods appropriate for a given health context.	Distinguish among the different quantitative and qualitative types of data (e.g., continuous, ordinal) that may be collected. Create new variables from given variables (e.g., tertiles of a continuous measure) and investigate whether transformations of a variable are needed to satisfy modeling assumptions or aid interpretation. Understand the importance of minimizing missing data.	Homework assignments and the group project will cover this learning objective.
3. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software.	Plan and implement a data analysis that is: (a) specific to the evaluation of a stated scientific/public health hypothesis; (b) appropriate for the type of data being analyzed; and (c) for which the data satisfy any assumptions. Understand the following methods for such analyses: multiple linear regression, logistic regression, Poisson regression, linear mixed models, generalized linear mixed models, Generalized Estimating Equations, Kaplan-Meier survival curves, and proportional hazards regression.	Homework assignments and the group project will cover this learning objective.
4. Interpret results of data analysis for public health research, policy or practice.	Interpret the results of your data analysis within the context of the stated hypothesis and the scientific/public health relevance.	Homework assignments and the group project will cover this learning objective.
18. Select communication strategies for different audiences and sectors. and 19. Communicate audience-appropriate public health content, both in writing and through oral presentation.	Be aware of the need to assess your audience's scientific and statistical backgrounds and public health research interests; practice tailoring your written communications to those backgrounds and interests.	The group project (both the short report and the poster) will cover this learning objective through the crafting of the language used within background, methods, results, public health/clinical relevance, and discussion.
11. Select methods to evaluate public health programs or policies. and 15. Evaluate policies for their impact on public health and health equity.	Decide upon research hypotheses for a dataset that involves public health programs or policies, design an analysis plan appropriate for the study design and the research questions, implement that analysis plan, interpret the results, and describe the entire study in a research summary report.	Groups may choose to use a dataset for their group project that involves public health programs or policies; the project would then cover these learning objectives.
21. Perform effectively on interprofessional teams.	Work as an interprofessional team to decide upon research questions, design an analysis plan appropriate for the study design and the research questions, implement that analysis plan, interpret the results, and describe the entire study in a research summary report.	Groups may self-select for their group project to include interprofessional members; the project would then cover this learning objective.