



POP HLTH 651

B M I 651

Advanced Regression Methods for Population Health

Advanced Regression Methods Pop Health

3 CREDITS

FALL 2024

Instructional Mode

Face to face

Location and Schedule

Tuesdays and Thursdays, 4:00-5:15pm

726 WARF

Canvas Course

<https://canvas.wisc.edu/courses/426488>

Course Instructor

Mary Ryan Baumann, PhD

Assistant Professor

Department of Population Health Sciences

Email: mary.ryan@wisc.edu

Office: 701 WARF

Instructor Office Hours: Th 10-11a (in-person); M 2-3p (Zoom)

Course Description

Extension of regression analysis to observational data with unequal variance, unequal sampling and propensity weights, clusters and longitudinal measurements, using different variance structures, mixed linear models, generalized linear models and GEE. Matrix notation will be introduced and underlying mathematical and statistical principles will be explained. Examples use data sets from ongoing population health research.

Requisites

POP HLTH/B M I 552.

Learning Objectives

This course provides an overview of modern statistical approaches to longitudinal and correlated data. It seeks to bridge the gap between statistical theory and real-world application by mixing methodological lectures with case studies using diverse types of data arising in health sciences research. The main emphasis is on the practical aspects of (clustered) data analysis.

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

1. Extend the knowledge of regression analysis beyond ordinary linear and logistic regression models

2. Understand the features of correlated data, their role in epidemiologic study, and their implications in drawing inference
3. State the assumptions underlying linear mixed effect, generalized linear mixed effect, generalized estimating equation, and survival regression models, recognize and address violations of those assumptions, and estimate and interpret regression models to answer epidemiologic and public health research questions in correlated and longitudinal data settings
4. Implement the inference procedures to solve real-world problems using statistical packages such as SAS and R
5. Interpret and present the analytic results to answer substantive questions

How Credit Hours are Met

The credit standard for this course is met by an expectation of a total of 135 hours of student engagement with the course learning activities (at least 45 hours per credit), which include readings, two in-person sessions per week, weekly homework exercises, one midterm exam, one data analysis project, and peer reviews as described in this syllabus.

Textbooks

There are no required textbooks for the course. Many of my course notes are based on the book

[Applied Longitudinal Analysis](#). 2011. Fitzmaurice GM, Laird NM, and Ware JH. John Wiley & Sons: Hoboken, New Jersey. ISBN: 9780470380277.

You can purchase a hard copy of the textbook from [University Book Store](#) or [Amazon](#). Electronic copies of the textbook can be accessed through [UW-Madison Library](#) or downloaded [directly from the publisher](#). An electronic copy of the textbook is also available on Canvas.

Supplemental Texts

There are several texts or portions of texts that I have found helpful. These books are not required, but I list them here as a resource if you are interested.

[Analysis of Longitudinal Data](#), 2nd edition. 2002. Diggle PJ, Heagerty P, Liang K, and Zeger SL. Oxford University Press: Oxford. ISBN: 9780198524847. Select chapters/sections of this book will be uploaded on Canvas for supplemental reading.

Chapters 11-15 of [Data Analysis Using Regression and Multilevel/Hierarchical Models](#). 2007. Gelman A and Hill J. Cambridge University Press: New York. ISBN: 9780521686891. Select chapters/sections of this book will be uploaded on Canvas for supplemental reading.

Chapter 9 of [Foundations of Linear and Generalized Linear Models](#). 2015. Agresti A. John Wiley & Sons: Hoboken, New Jersey. ISBN: 978-1-118-73030-0. [Ebook available through UW Library](#), as well as on Canvas.

Chapters 7-11 of [Beyond Multiple Linear Regression: Applied Generalized Linear Models and Multilevel Models in R](#). 2021. Roback P and Legler J. CRC Press. ISBN: 9781439885383. [Ebook available for free online](#).

(Tentative) Course Schedule

There are 13 modules in the course.

Module/Theme	Date/Topic	Suggested Readings	Assignments assigned	Assignments due	Notes
Week 0	Th 9/5: Review of (independent) regression methods	Fitzmaurice p. 1-16	- <i>Small assignment 1: Introduce yourself</i> – due 9/10 - <i>Homework 0</i> – due 9/12 - <i>Milestone 1: Determine final project context/data</i> – due 9/26		
Week 1 <i>Introduction to correlated data</i>	T 9/10: Correlated non-longitudinal studies	Gelman p. 237-248; <u>Beyond MLR Ch. 7</u>	- <i>Homework 1</i> – due 9/17	- <i>Small assignment 1: Introduce yourself</i> – due 9/10	
	Th 9/12: Hands-on analysis tutorial			- <i>Homework 0</i> due @ 4pm	
Week 2 <i>Conditional models for correlated data</i>	T 9/17: Linear mixed effect models (LMEs)	Agresti p. 286-296; Gelman p. 251-276; <u>Beyond MLR Ch. 8</u> ; McNeish & Kelley (2019); Clark & Linzer (2014)	- <i>Homework 2</i> – due 9/24	- <i>Homework 1</i> due @ 4pm	
	Th 9/19: Hands-on analysis tutorial				
Week 3 <i>Non-normal responses, correlated data, and conditional models</i>	T 9/24: Generalized linear mixed effect models (GLMMs)	<u>Beyond MLR Ch. 5</u> ; <u>Beyond MLR Ch. 6</u> ; Agresti p. 307-310	- <i>Homework 3</i> – due 10/1 - <i>Milestone 2: Brainstorm final project</i>	- <i>Homework 2</i> due @ 4pm	

			<i>potential research questions – due 10/10</i>		
	Th 9/26: Hands on analysis tutorial			- Milestone 1: Final project broad context due @ 11:59pm	
Week 4 <i>Marginal models for correlated data</i>	T 10/1: Generalized Estimating Equations (GEEs)	Agresti p. 314-319; Fitzmaurice p. 359-361; Hubbard et al. (2010)	- Homework 4 – due 10/8 - Small assignment 2: GEE vs GLMM (pt 1) – due at midterm (10/17)	- Homework 3 due @ 4pm	
	Th 10/3: Hands-on analysis tutorial				
Week 5 <i>Modeling correlated count data & correlated data model diagnostics (pt 1)</i>	T 10/8: Correlated count data and diagnostics	<u>Beyond MLR Ch. 4;</u>	- Homework 5 – due 10/15	- Homework 4 due @ 4pm	
	Th 10/10: Hands-on analysis tutorial		- Milestone 3: Settle on final project primary research question – due 10/24	- Milestone 2: Final project potential research questions due @ 11:59pm	
Week 6 <i>Longitudinal studies</i>	T 10/15: Correlated longitudinal studies	Fitzmaurice p. 19-44		- Homework 5 due @ 4pm	
	Th 10/17: Midterm			- Small assignment 2: GEE vs GLMM (pt 1) due @ 4pm	
Week 7 <i>Modeling covariance</i>	T 10/22	Diggle p. 33-53;	- Homework 6 – due 10/29		

		Fitzmaurice p. 165-186			
	Th 10/24: Hands-on analysis tutorial		- Milestone 4: Final project proposed model – due 11/7	- Milestone 3: Final project specific research question due @ 11:59pm	
Week 8 <i>Conditional & marginal models for longitudinal correlation</i>	T 10/29: GLMMs and GEEs for longitudinal/ ordered data	Agresti p. 299-302		- Homework 6 due @ 4pm	
	Th 10/31: Hands-on analysis tutorial		- Homework 7 – due 11/12		
Week 9 <i>Longitudinal model diagnostics</i>	T 11/5: GLMMs and GEEs (cont), and diagnostics	Fitzmaurice pg. 189- 209, pg. 353-361, & pg. 381-385			Election day!
	Th 11/7: Hands- on analysis tutorial		- Milestone 5: Final project Table 1 – due 11/21	- Milestone 4: Final project proposed model due @ 11:59pm	
Week 10 <i>Introduction to time-to-event data</i>	T 11/12: Time- to-event data and studies	Kleinbaum pg. 1-82; Moore pg. 1-52		- Homework 7 due @ 4pm	
	Th 11/14: Kaplan-Meier estimator and log-rank test				

Week 11 <i>Analysis of time-to-event (pt 1)</i>	T 11/19: Hands-on analysis tutorial		- Homework 8 – due 12/3 - Final papers – due 12/16 - Project presentation sign up – due 12/3		
	Th 11/21: Cox proportional hazards regression	Kleinbaum pg. 97-145; Moore pg. 55-71		- Milestone 5: Final project Table 1 due @ 11:59pm	
Week 12 <i>Analysis of time-to-event analysis (pt 2)</i>	T 11/26: Hands-on analysis tutorial				
	Th 11/28				Thanksgiving break – no class
Week 13 <i>Special topics & Final project presentations</i>	T 12/3: Other multi-level models/ confounding issues			- Homework 8 due @ 4pm - Project presentation sign up due @ 5:15pm	
	Th 12/5: Project presentations				
Week 14 <i>Final project presentations</i>	T 12/10: Project presentations				
	Th 12/12:				Study day – no class
Finals week	M 12/16			- Final papers due by 11:59pm	

Clean (unannotated) PDF versions of the lecture slides can be found on Canvas; HTML versions of the slides are hosted at a generic web address: <https://maryryan.github.io/PHS-651-slides/PHS-651-WeekNum/slides-WeekNum>. For example, the slides for week 2 lectures can be found at: <https://maryryan.github.io/PHS-651-slides/PHS-651-2/slides-2>.

Course Requirements and Evaluation

Your grade will be based on completion of the weekly homework/small assignments (40% of your grade), a midterm exam (15% of your grade), and a final data analysis project (45% of your grade).

Homework/Small Assignments (40%)

We will have weekly homework assignments throughout the course. The assignment will be released after the Tuesday session (5:15pm), and will be due before the following Tuesday session (4pm). Homework is intended to give you a chance to practice concepts we learned in class.

There will also be several “small assignments” that will be assigned over the course of several weeks; their assignment and due dates are indicated in the course schedule in the syllabus. The purpose of these small assignments is to review several weeks’ worth of topics, and compare and synthesize the concepts learned.

For both weekly homework and small assignments, you may collaborate with your peers but you will each turn in the assignment individually and the end product is expected to be your own work and your own words.

Midterm Exam (15%)

This course will include a written midterm exam to be held during class time on Thursday 10/17. It will cover topics from the beginning of class through to the end of Week 5 (no longitudinal studies). The exam will be closed-book and closed-note. The purpose of the exam is to assess your understanding of the topics covered through Week 5, as well as your ability to make and justify analytic choices. ***If you have a conflict with the exam date, please alert the instructor as soon as possible.***

Final Data Analysis Project (45%)

Throughout this course you will work on a semi-open-ended data analysis project. The goal is for you to appropriately use any of the knowledge and skills you gain in this class to investigate a research question of your choosing. In this class we believe that null results are valid results – it is not the objective to game your analysis or your questions until you have a statistically significant coefficient. You may choose between several provided datasets; you may also use a dataset you have found, but you must schedule a brief appointment to discuss and have it approved by me. Longitudinal data may be “converted” to cross-sectional data if the appropriate research question is of interest and the analysis uses methods from this course.

Intermediate deadlines will be set throughout the semester to keep you on track and to provide you with formal feedback (collectively 10% overall grade). You may alter these “milestones” after the interim due date and prior to turning in the final project if you wish – these interim deadlines are not meant to “lock” you in to any particular analysis aspect but it is advised you consult with me on major changes (i.e., dataset, primary research question). You will also be assigned to “feedback group” – a group of your peers with whom you’ll have multiple opportunities to discuss and solicit feedback on these milestones before they’re due. Feedback groups aim to guarantee everyone with some continuing level of informal brainstorming/feedback peer support through the semester; they do not make this a group project.

The final product will consist of one 8-10-minute oral presentation (10% overall grade) and one written report (25% overall grade). You will be assigned 3 oral presentations to provide formal written feedback on, though you may (and are encouraged to!) provide informal feedback for any presentation. Oral presentations will not be recorded so synchronous attendance during presentation week is mandatory. The written report has a 10-page maximum (excluding appendices; no minimum), and will consist of the following sections: introduction/background, methods (data source, statistical methods), results, and

discussion/limitations. You may use appendices for any additional tables/plots/analyses you don't feel appropriate for the main report, but it is not a required section. Reports should be written in manuscript-type style (i.e., to the general scientific community), with additional description justifying the appropriateness of the statistical methods used. You may discuss your project with me or your peers throughout the semester, but the work you turn in must be your own.

Oral presentations will be graded based on 2 main criteria:

1. Scientific approach (40%)
 - a. Did you appropriately state your research question(s)?
 - b. Did you provide a concise overview of the statistical methods used to answer the question(s) and brief justification for why they were used?
 - c. Did you appropriately interpret your results?
 - d. Did you discuss any limitations?
2. Peer feedback (20% \times 3 [60% total])

Written reports will be graded based on 3 main criteria:

1. Scientific approach (30%)
 - a. Did you succinctly state your research question(s)?
 - b. Did you consider/investigate problems in the sampling that might materially affect the results?
 - c. Did you thoroughly explain and justify the rationale behind your modeling choices?
 - d. Did you choose appropriate models to answer the scientific questions?
2. Statistical approach (45%)
 - a. Were the methods chosen appropriate for the data at hand? Were any key assumptions violated?
 - b. What did you do to address violations? Did your 'fixes' material change your conclusions?
3. Written report (25%)
 - a. Were your findings well-documented in a succinct manner (not too verbose, not missing necessary details)?
 - b. Was the report written to an appropriately broad audience?
 - c. Overall, did you clearly explain your choices, stick to your research question(s), and provide precise interpretations of your statistical results?
 - d. Were your results presented in an appropriate manner? (no raw code output!)
 - e. Did you adequately address/acknowledge limitations?
 - f. Did you incorporate peer feedback from oral presentation?

You should incorporate *relevant* pieces of statistical output into your report in *self-contained figures and tables*. You should also include **relevant code and annotated output (be selective)** as an attached .zip to facilitate grading, but your report itself should not directly depend on either the code or the output.

Remember *all models are wrong; the practical question is how wrong do they have to be to not be useful*. Most importantly, remember that the goal of any statistical analysis is to **learn about the actual relationship between the variables**; it is **not to find a statistically significant result**.

Software

You are welcome to use any statistical software for the exercises and data analysis projects. I will use SAS and R for course instruction, and the TA and I will provide support for these languages. If you choose to use other software, instructors will provide as much assistance as they can, but support cannot be guaranteed.

SAS Set-up

There are several ways to access SAS software at no charge. SAS OnDemand for Academics is the recommended option. It is a cloud-based system accessed using a web browser.

You are invited to enroll in the SAS OnDemand for Academics course (Advanced Regression Methods for Population Health, Fall 2024) associated with this course.

As a first step, please create your account for SAS OnDemand for Academics. To register, visit <https://odamid.oda.sas.com> and click on *Register for an account*.

After you have successfully created your account, follow these steps:

1. Sign on the Control Center at <https://odamid.oda.sas.com>.
2. Look for the *Enroll in a course* link in the "Enrollments" section near the bottom of the page. Click this link to start the enrollment.
3. Enter the course code: 32469ccd-38ac-4fc7-a3d1-547b1758180d
4. Submit the form.
5. Confirm that this is the correct course and then click the button to finish enrolling.

I plan to teach using the following software: SAS Studio.

I have also uploaded data for us to use in our course. You will be able to access that data using a LIBNAME or FILENAME statement with this path: ~/my_shared_file_links/u63698550

For more information about SAS OnDemand for Academics, including step-by-step registration instructions, visit the following site: <http://support.sas.com/ondemand>.

The regular edition of SAS is licensed for installation on any UW-owned device or a device owned by (currently employed) UW-Madison faculty and staff (but not students). SAS is only available for Windows and Linux operating systems. It can be downloaded from the [UW-Madison Campus Software Library](#). (Note that these files are very large. They will take a long time to download even on a wired connection. It is strongly recommended that you not use a wireless connection to download these files.)

If you have not taken Pop Hlth/BMI 451 Introduction to SAS for Population Health, you should either review the [Introduction to SAS class notes](#) from the UCLA Institute for Digital Research and Education or enroll in the [SAS Programming 1: Essentials](#) course offered by SAS (the self-paced, e-Learning version of the course is free).

Communications Policy

1. I will send a weekly email on Friday evenings; it will also be posted as an announcement on the course website. This email/announcement will briefly list the topic, activities and assignments for the next week in class.

2. There will be a discussion forum [on Piazza](#) for student questions. It is my expectation that most questions can (and will) be answered by your fellow students.
3. I will be available via email (mary.ryan@wisc.edu) Monday – Friday 8a-6p. You can expect a response to your email within 24 hours, excluding weekends and holidays. I may be checking my email or reply outside of these hours, but that should not be the expectation.
4. I will hold in-person office hours on Thursdays from 10-11a, and Zoom office hours on Mondays from 2-3p. You can stop by to discuss any topic (related to class or not), but class topics will be given priority if multiple people are present. If you cannot make office hours or would like to discuss something one-on-one, please schedule an appointment with me via my calendly: <https://calendly.com/mary-m-ryan/>
5. I maintain an open-door policy, and I am happy to discuss statistical questions, related or unrelated to the course, with you at any opportunity. If you're in the WARF building, please feel free to drop by my office anytime. If you would like to guarantee an uninterrupted one-on-one (or many-on-one) meeting, I encourage you to schedule an appointment for an in-person meeting (or Zoom call) via my calendly: <https://calendly.com/mary-m-ryan/>
6. The TA will hold office hours with times set weekly. She will send out a poll each week to determine the office hours for the following week. You may also email her to schedule time to meet outside of the set hours.

Attendance Policy

I do not take attendance for this course. Due to varying work schedules, personal responsibilities, or accessibility issues students may have, attending an in-person lecture or discussion may not always be feasible. All in-person lectures will also have a synchronous Zoom-in option. I strongly encourage students to attend lectures in-person if they can, or live on Zoom if in-person isn't possible, so you can fully take advantage of in-class exercises, ask questions, and participate in live discussions. In addition, scheduled lecture and discussion sessions are time set aside each week for you to complete assigned work. Regardless, all in-person lectures will be recorded and later posted to the Canvas website for use in the event of schedule conflicts or studying purposes. All students are responsible for keeping track of assignments and watching video content, regardless of whether they attend live sessions.

Homework/Assignment Honesty Policy

Although you are encouraged to study discuss homework with other students, you are each expected to do your own work. If you are assigned a problem that has an answer in the student solutions manual or the back of the book, you are expected to write the solution in your own words, not copy it from the book. If solutions to an assignment exist on a website somewhere (such as, but not limited to, Chegg, Course Hero, or Reddit) it is a violation of the honesty policy to directly copy them. You also may not upload assignments to sharing websites such as, but not limited to, Chegg, Course Hero, or Reddit. Students found to be in violation of academic integrity will receive a 0 for the assignment and will be reported to the Office of Academic Integrity & Student Conduct.

Late Work/Extensions Policy

There is a 2-hour grace period for late work; after this grace period, late work will be docked 50% for each day late (i.e., work that is turned in 3 hours late will automatically have 50% taken off). If you need extra time for an assignment, contact me via email (mary.ryan@wisc.edu) **before the due date** to

discuss an extension. Extensions will be considered on a case-by-case basis. All work received, no matter when it is turned in, will receive feedback.

AI Statement

The use of artificial intelligence (AI) tools and applications (including, but not limited to, ChatGPT, Copilot, DALL-E, and others) for course assignments and assessments does not support the learning objectives of this course and is prohibited. Using them in any way for this course is a violation of the course's expectations and will be addressed through UW–Madison's [academic misconduct policy](#), specifically UWS 14.03(1)b (b) Uses unauthorized materials or fabricated data in any academic exercise.

Teaching & Learning Data Transparency Statement

The privacy and security of faculty, staff and students' personal information is a top priority for UW-Madison. The university carefully evaluates and vets all campus-supported digital tools used to support teaching and learning, to help support success through [learning analytics](#), and to enable proctoring capabilities. View the university's full [teaching and learning data transparency statement](#).

Privacy of Student Records & the Use of Audio Recorded Lectures Statement

View [more information about FERPA](#).

Lecture materials and recordings for this course are protected intellectual property at UW-Madison. Students in this course may use the materials and recordings for their personal use related to participation in this class. Students may also take notes solely for their personal use. If a lecture is not already recorded, you are not authorized to record my lectures without my permission unless you are considered by the university to be a qualified student with a disability requiring accommodation. [Regent Policy Document 4-1] Students may not copy or have lecture materials and recordings outside of class, including posting on internet sites or selling to commercial entities. Students are also prohibited from providing or selling their personal notes to anyone else or being paid for taking notes by any person or commercial firm without the instructor's express written permission. Unauthorized use of these copyrighted lecture materials and recordings constitutes copyright infringement and may be addressed under the university's policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

Digital Course Evaluation (AEFIS)

UW-Madison uses a digital course evaluation survey tool called [AEFIS](#). For this course, you will receive an official email two weeks prior to the end of the semester, notifying you that your course evaluation is available. In the email you will receive a link to log into the course evaluation with your NetID. Evaluations are anonymous. Your participation is an integral component of this course, and your feedback is important to me. I strongly encourage you to participate in the course evaluation.

Diversity & Inclusion Statement

[Diversity](#) is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals. The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world.

Academic Integrity Statement

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion.

Accommodations for Students with Disabilities Statement

The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform the instructor of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. The instructor will work either directly with the student or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. Students may also manage accommodation requests via [McBurney Connect](#). (See: [McBurney Disability Resource Center](#))