STAT 437: Statistical Methods for Life History Analysis

Winter 2022

An online version of this course outline is available at: https://outline.uwaterloo.ca/view/2033.

Instructor:	Dylan Spicker	**Lectures:**	Online!
Email:	dylan.spicker@uwaterloo.ca	**Office Hours:**	Online By Appointment
Pronouns:	They/Them	**Tutorial:**	Online (Timing TBA)

This course covers a wide range of topics pertaining to the statistical analysis of data from prospective studies. The major areas of coverage for this course are i) methods for the analysis of longitudinal data arising from repeated measurements of individuals and ii) the analysis of life history data. For longitudinal data, three general classes of methods will be discussed, including robust marginal models based on estimating equations, transitional models and mixed effects models. Continuous and discrete outcomes will be treated. The material on event history analysis will primarily involve an introduction to the analysis of survival data with some introduction to the modeling of censoring, truncation and more complex sampling processes. Studies will be discussed from the epidemiological literature and other sources in the public domain. Data analysis will be carried out using statistical software (e.g., R or SAS).

Territorial Acknowledgement

The University of Waterloo acknowledges that much of our work takes place on the traditional territory of the Neutral, Anishinaabeg and Haudenosaunee peoples. Our main campus is situated on the Haldimand Tract, the land granted to the Six Nations that includes six miles on each side of the Grand River.

I would like to acknowledge that (since I am in Ottawa) I am on the traditional unceded territory of the Anishnaabeg People. I recognize that we may all be in different locations, and that you may be present in a different, traditional Indigenous territory. If so, I encourage you to take a moment to reflect on that and acknowledge it.

Learning Environment: In this course I will endeavour to create a welcoming environment for all students, so as to best facilitate everyone's learning. I invite you, if you are comfortable, to share the name and pronouns that you would like us to use in this space (whether that is to me directly over email or to your classmates on the platforms on which we interact). I also invite you to share with me any other information regarding your circumstances which is important for me to take into consideration during this course. If I can be of support to you, or help put you in touch with resources you require, I will.

Course Website: Please visit https://learn.uwaterloo.ca/ for all course related material. It may be worth considering your email and notification settings as Learn will be the primary means of communication for course updates throughout the term.

Prerequisites: This course will rely heavily on material from STAT 331 and STAT 431. I will provide review of this material in an attempt to keep it self-contained, but the review will be fairly brief, and will assume some familiarity.

Office Hours: By online appointment. There is a booking link provided on Learn, alongside the full booking policy. If there is a need outside of the allotted times, please email me, and we can discuss further.

Discussion Forum: Seeing as this course is online, we will use an online location for us to discuss the course material, and for you to ask questions of me and each other. I propose that this takes place on Teams, which you should be automatically enrolled in, and is linked on the Learn. I am open to other options and would prefer to hear from you as to what has worked best in the past (for instance, we could consider Discord or Slack, instead).

Main References: None of these materials are required. All of the required content in the course will be covered in the lecture videos and in supplemental notes (as are required) posted to the course website. These are a selection of potentially useful textbooks for learning the material. Other books are likely acceptable and encouraged if they help you approach the material! **These books are all available through the library on course reserves.**

- Fitzmaurice, Laird, and Ware. *Applied Longitudinal Analysis.* 2011. (John Wiley and Sons).
- Fitzmaurice, Davidian, Verbeke, and Molenberghs. *Longitudinal Data Analysis.* 2008. (Chapman & Hall/CRC Handbooks of Modern Statistical Methods).
- Diggle, Heagerty, Liang, and Zeger. *Analysis of Longitudinal Data.* 2002. (Oxford University Press).
- Klein, and Moeschberger. *Survival Analysis: Techniques for Censored and Truncated Data.* 2003. (Springer).
- Singer JD and Willett JB. Applied longitudinal data analysis: Modeling change and event occurrence. Oxford University Press, 2003.

Objectives: Primarily, this course is designed to give you experience (conceptual, applied, and theoretical) with the analysis of longitudinal and survival data – topics which have important implications for scientific inquiry. I recognize that each of you is entering this class with different goals, expectations, and desires, and it is my intention to help you achieve those outcomes, while exposing you to these statistical concepts. You will be encouraged (early in the term) to provide some guidance as to your goals, plans, and interests. This will help me shape your experience in this course in service of those outcomes. In exchange, I hope that you take the time to approach the course material with interest and critically engage with the ideas presented. By the end of this course students should be able to:

- Describe the features of longitudinal data, the mechanisms through which they arise, and the scientific questions which require their use (particularly as compared to cross-sectional data).
- Understand marginal models for both continuous and discrete outcomes.
- Understand and explain generalized estimating equations (GEE), as they pertain to marginal models.
- Understand mixed effect models for both continuous and (time permitting) discrete outcomes.
- Understand transitional models for categorical data.
- Describe, interpret, and communicate results from the various estimation techniques for longitudinal models, contrasting interpretations across models.
- Describe the features of time to event data, particularly with respect to censoring and truncation mechanisms.
- Understand and describe the key quantities of interest for survival analysis (namely, the survival and hazard functions).
- Understand parametric estimation for both discrete and continuous time to event models.
- (Time permitting) Understand nonparametric estimation of the survival function (through Kaplan-Meier curves).

Note: the topics covered in this course will be explored conceptually, theoretically, and with application. This will include using mathematical derivations to explore ideas, fitting and interpreting models through statistical software (e.g., R), and with an exploration of existing scientific literature.

Tentative Course Schedule
Note, the following course schedule is subject to change before the start of the course.

Week	Dates	Course Deadlines	Suggested Video Material
1	Jan. 5 – Jan. 7		
2	Jan. 10 – Jan. 14		
3	Jan. 17 – Jan. 21	**January 21:** Problem Set #1 Due	
4	Jan. 24 – Jan. 28		
5	Jan. 31 – Feb. 4	**February 4:** Problem Set #1 Resubmit†	
6	Feb. 7 – Feb. 11	**February 11:** Problem Set #2 Due	
7	Feb. 14 – Feb. 18		
–	Feb. 21 – Feb 25	**Reading Break:** There is no course content to be covered during your reading break. Please take the time for yourself!	
8	Feb. 28 – Mar. 4	**March 4:** Problem Set #2 (Resubmit)†	
9	Mar. 7 – Mar. 11	At-home term test to be administered during this week.	
10	Mar. 14 – Mar. 18	**March 18:** Problem Set #3 Due	
11	Mar. 21 – Mar. 25		
12	Mar. 28 – Apr. 1; Apr. 4 – Apr. 5	**April 1:** Problem Set #3 (Resubmit)†; **April 5:** Paper Critique Due	

Items marked with a † are optional.

Course Schedule: The course is run asynchronously. The lecture videos are broken down by topics, rather than by weekly sessions, in an attempt to make them more digestible in small chunks. When relevant, topics are further divided into conceptual, theoretical, and applied videos.

- **Conceptual Videos:** describe the specific details of the relevant topic, and encourage an intuitive understanding. These are necessary videos for covering the content, and encourage deeper statistical literacy. In these videos, definitions and important results are presented, models are described, and so forth, however, mathematical proofs and applied examples are (mostly) skipped over.
- **Theoretical Videos:** cover the deeper mathematical proofs and derivations for the relevant concepts outlined in the conceptual lectures. These proofs and derivations are walked through in detail, and where appropriate, you are encouraged to try to discover the details for yourself.

- **Applied Videos:** look at the use of the concepts to answer scientific and applied statistical questions. These videos will most often take the form of analyzing real data sets using R, both demonstrating how to use the statistical tools to fit the models we are discussing, and how to perform diagnostics on the models, interpret the parameter estimates, and communicate these results in a scientific setting. The data used in these videos are made available on the course website.

All videos are made available both on YouTube (links on Learn) and on Learn directly, and can be watched at your chosen pace. The videos which are relevant for given assignments will be clearly communicated on each assignment (on Learn) to ensure that you are not falling behind on the material. Each week I will communicate what specific progress should be made based on how the term is going. Any changes to these videos (for instance, due to correcting mistakes) will be announced throughout the term, as is necessary.

Tutorials: The course has an online tutorial slot. The specific use of this time will be discussed as a class to determine what will best facilitate your learning.

Assessment Breakdown: Grades in the course will be determined based on: Three problem-set assignments (15% each, 45% total); A take-home term test (15%); A paper critique/review (10%); and a Final cumulative project (30%).

Assignments: The purpose of assignments is to give you practice with the course concepts, and to provide a self-assessment for the concepts which are understood and those which are not. Assignments are to be submitted through Crowdmark, and you will receive a personal link via email well before the due date.

- **Assignment Grading:** Each assignment has two corresponding deadlines, a submission date and a later re-submission date. After the submission date, your assignment will be graded, and high-level feedback will be given, in advance of the re-submission date. Following this, you will have the opportunity to (if desired), using the feedback, re-submit your assignment with corrections made, and a brief written summary explaining the initial error (i.e., what was the misunderstanding). I will then mark the re-submitted papers, allowing for (up to) 50% of the lost marks to be returned. At this point, solutions to the assignment questions will be released on Learn. Note, if you do not re-submit your assignment, the grade originally presented will serve as the grade on the assignment. Full details will be provided alongside the first assignment.

- **Late or Missed Assignments:** If you do not submit before the submission date, without a prior extension, you will still be able to submit your work by the re-submission date (allowing you to earn up to 50% on the assessment). If you do not submit by the re-submission date, your work will not be accepted, and you will receive a 0% on the corresponding assignment.

- **Rules for Group Work:** Learning statistics, particularly online, is more effectively done with peers to assist and challenge your understanding. You are encouraged to discuss problem sets with your classmates, but each student must submit their own unique work (solutions, code, results, etc.). Further, any students who discuss solutions with their peers should indicate this on their submission, to avoid being flagged for academic integrity violations.

- **Retrieving Marked Assignments:** A link to your graded submission (and re-submission, if applicable) will be sent to you via email.

Term Test: The term test (tentatively the week of March 7th) will be a take-home, open-book exam. You will be expected to complete the material by yourself (without input from your peers). You will be given several days to complete the midterm, though the intention is not that it will take this full time: instead, the time is intended to give you flexibility for when you complete the work. Details regarding the precise coverage will be provided during the term, based on the progression through the material. The midterm exam will not have a re-submission date.

Paper Critique: A significant component of this course, and indeed any advanced statistics course, is developing statistical literacy which will assist in the interpretation of scientific results. This will be a focus throughout the term, and late in the term you will be asked to submit a review/critique/interpretation of a paper which is relevant to the material covered in the course. You will get to choose the specific form that this will take (e.g., a written response, an academic poster, a presentation video, etc.) and more details will be released during the term. The paper critique will not have a re-submission date.

Final Project: You will be assigned a cumulative project which will ask you to synthesize the topics covered throughout this course with a real-world application of the methods, or by engaging with extended theoretical concepts beyond those covered in the course material. Just as with the paper critique, further specific guidance will be provided once the term has started, but the idea with the project will be to give you significant choice and freedom, to hopefully tailor the project to your goals and objectives with this course and beyond. Tentatively, the 30% will be divided into two components: a submitted component, worth 20% of the final grade, and an informal interview/discussion (with me on Teams or Zoom) which focuses primarily on your rationale for choices made in the project and how that relates to your understanding of the course concepts, worth 10% of the final grade. The final project will not have a re-submission date.

Tentative Assessment Schedule (Subject to Change)

	Submission Deadline	**Re-Submission Deadline**
Problem Set #1	Friday, January 21st	Friday, February 4th
Problem Set #2	Friday, February 11th	Friday, March 4th
Problem Set #3	Friday, March 18th	Friday, April 1
Term Test	Week of March 7th	Not Applicable
Paper Review Critique	Tuesday, April 5	Not Applicable
Final Project	Monday, April 25	Not Applicable

Course Policies:

- **Attendance:** The course is being offered online, asynchronously, in an attempt to best accommodate everyone's schedules. With that said, our weekly tutorial sessions (if they occur) will occur live and will not be recorded. This policy is to encourage students to ask questions and participate freely, without concern over the recording. The exact use of the tutorial sessions will be decided upon as a class, and I hope that your desire to attend will be rooted in their utility. With that said, there is no attendance grade for the course and students are responsible for any content that may be missed. If scheduling the tutorials becomes an issue, we can have a discussion about alternative arrangements to ensure that every student has equitable access to the material.
- **Requests for Re-grading:** If, after reviewing the posted solutions and/or feedback, you feel your assignment/test/project has an incorrect grade, you are entitled to request a re-grade. Please email me your request, with a brief explanation specifying which question/component is to be re-graded, and your rationale for the re-grade request. Requests should be made within seven days of the day that the assessment was returned (or seven days following the solutions, if those are delayed).
- **Accommodations for Extenuating Circumstances:** If you require an accommodation for a date or deadline due to extenuating circumstances (e.g., illness) please reach out to me as early as possible.

In certain situations, in accordance with university policy, you may be required to submit verification of the extenuating circumstance. In the event of extenuating circumstances, the weight of the component will either be moved to the final project or an extension may be granted.

A Note on Grades: Grading is intended to be a mechanism for communicating your understanding of the material, both to you (throughout the term) and to others (after the term). It is not my intention to use grades as a means of motivating your progression through the course. Instead, I have designed the assessments in the course to serve as useful for you, directly, following the course completion (for instance, as portfolio items for possible data science jobs or examples of research for graduate school applications). I encourage you to reach out to me if you have specific goals which we may be able to achieve through the use of the assessed materials in the course, allowing for them to serve as more than a vehicle towards a numeric grade. At the same time, I recognize that grades can be a source of anxiety for students. I have attempted to create a grading scheme which is fair and throughout the term I will communicate exact standards for work. I encourage you to reach out and discuss with me if there are any ways we can ensure that stress over grades does not prohibit learning in this course.

University Policies

- **Academic integrity**: In order to maintain a culture of academic integrity, members of the University of Waterloo community are expected to promote honesty, trust, fairness, respect and responsibility. [Check the Office of Academic Integrity for more information.]

- **Grievance**: A student who believes that a decision affecting some aspect of their university life has been unfair or unreasonable may have grounds for initiating a grievance. Read Policy 70, Student Petitions and Grievances, Section 4. When in doubt, please be certain to contact the department's administrative assistant who will provide further assistance.

- **Discipline**: A student is expected to know what constitutes academic integrity to avoid committing an academic offence, and to take responsibility for their actions. [Check the Office of Academic Integrity for more information.] A student who is unsure whether an action constitutes an offence, or who needs help in learning how to avoid offences (e.g., plagiarism, cheating) or about "rules" for group work/collaboration should seek guidance from the course instructor, academic advisor, or the undergraduate associate dean. For information on categories of offences and types of penalties, students should refer to Policy 71, Student Discipline. For typical penalties, check Guidelines for the Assessment of Penalties.

- **Appeals**: A decision made or penalty imposed under Policy 70, Student Petitions and Grievances (other than a petition) or Policy 71, Student Discipline may be appealed if there is a ground. A student who believes they have a ground for an appeal should refer to Policy 72, Student Appeals.

- **Note for students with disabilities**: AccessAbility Services, located in Needles Hall, Room 1401, collaborates with all academic departments to arrange appropriate accommodations for students with disabilities without compromising the academic integrity of the curriculum. If you require academic accommodations to lessen the impact of your disability, please register with AccessAbility Services at the beginning of each academic term.

- **Turnitin.com**: Text matching software (Turnitin®) may be used to screen assignments in this course. Turnitin® is used to verify that all materials and sources in assignments are documented. Students' submissions are stored on a U.S. server, therefore students must be given an alternative (e.g., scaffolded assignment or annotated bibliography), if they are concerned about their privacy and/or security. Students will be given due notice, in the first week of the term and/or at the time assignment details are provided, about arrangements and alternatives for the use of Turnitin in this course. It is the responsibility of the student to notify the instructor if they, in the first week of term or at the time assignment details are provided, wish to submit alternate assignment.