BIOS 6312
Modern Regression Analysis
Spring 2021

Updated: 03/12/2021

Instructor: Andrew Spieker
Email: andrew.spieker@vumc.org
Office: 2525 WEA, 1137

Instructor office hour: Tuesday 12:00 PM - 1:00 PM

Class sessions: Tuesday 10:30 - 12:00 PM
Thursday 10:30 - 12:00 PM

Lab section: Monday, 1:00 - 2:00 PM

TAs: Jamie Joseph (jamie.g.joseph@vanderbilt.edu)
Caroline Birdrow (caroline.i.birdrow@vanderbilt.edu)

TA office hour: Wednesday, 3:45 - 4:45 PM

Course website: `https://andrewspieker.com/b6312`

Course description: This course covers modern regression analysis from an applied and methodological perspective. Specific topics to be covered include regression modeling for continuous, binary, nominal, ordinal, count, and time-to-event outcomes. This course will also cover more advanced topics such as basis and smoothing splines, model validation, prediction, and longitudinal data. Emphasis will be placed on approach, strategy, and interpretation of results. Additionally, this course will introduce approaches to handle challenges encountered in the real world (e.g., missing data). Theoretical principles will be demonstrated with real-world examples from biomedical studies.

Pre/co-requisites: Prior satisfactory completion of material covered in BIOS 6311 (or equivalent) is presumed (this includes descriptive statistics, point and interval estimation, one-sample and two-sample statistical inference). The lab section, BIOST 6312L, is required and must be taken concurrently. This course relies on the statistical software program Stata; prior experience specifically with Stata is not required, but some familiarity with statistical programming is expected.

Required notes: The lecture slides have been posted on the course website and are the major source of required material for this course. There is no required textbook.

References: The (non-required) texts below can serve as solid reference material, but where they disagree with lecture material, lecture material takes precedence.

- Vittinghoff, Glidden, Shiboski, and McCulloch. Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models, 2nd edition. Springer.
- Kleinbaum, Kupper, Muller, and Nizam. Applied Regression Analysis and Other Multivariable Methods, 3rd ed. Duxbury Press, 1998
- Kleinbaum and Klein. Logistic Regression: A Self-Learning Text (Third Edition) Springer, New York 2010.
- Kleinbaum and Klein. Survival Analysis: A Self-Learning Text (Third Edition) Springer, New York, 2012.

Learning objectives: By the end of the course, students should ordinarily expect to meet the following objectives:

- Understand how statistical methodology is motivated by medical problems.
- Become familiar with statistical concepts including exploratory data analysis, estimation, testing, and prediction.
- Select/implement regression methods to answer scientific questions, including those appropriate for continuous, binary, nominal, ordinal, count, and time-to-event outcomes.
- Properly specify transformations, categorization, and interaction terms in regression models and articulate proper parameter interpretation.
- State the statistical assumptions that are the basis for the conclusions of your analysis, and use diagnostic procedures to determine if assumptions are violated.
- Develop data analytic skills including familiarity with statistical software.
- Develop writing skills needed to communicate the results of a data analysis to a statistically naive reader.

I will emphasize the principles of the methodology rather than the mathematical underpinnings. This course is targeted to people who want to be able to read biomedical literature critically, and will be regularly analyzing data as part of their research.

Expectations and policies:

Expectations you can have of me: You can expect exams and problem sets to be graded in a timely fashion. You can expect me to be responsive to your questions about course material and/or grading. If you don't receive a response within a day, don't hesitate to e-mail again.

Recorded lectures: I plan to record lectures; I must emphasize that recordings are not a substitute for attendance. If I find attendance to be dropping, I reserve the right to halt recordings.

Collaboration: You are encouraged to work together on problem sets, with the caveat that your write-up should be in your own words. Exams are an individual effort.

Grading: If you believe there was a grading error, please direct your concern to Andrew.

Academic honesty: Students are encouraged to familiarize themselves with academic honesty policies. If you hand in an assignment not written in your own words, I cannot give you credit.

Extra help: Some concepts we cover in this course may be challenging. If you cannot attend my office hours, I strongly encourage you to make an arrangement with me to get help.

Late work: Work should be turned in on time, but I do understand that life happens. Late work must be approved by Andrew, in advance when possible.

Course evaluations: Please complete end-of-semester course evaluations. I read all comments closely and take them seriously. Changes from the previous year are driven by prior students' thoughtful feedback. Comments about what works well and specific, constructive suggestions for improvements provide a mechanism for me to improve the course in future years.

Voicing concerns: Keeping the above in mind, please do not feel obligated to wait until the end of the semester evaluations for your voice to be heard. If you have concerns about the material, its presentation, or how you're being evaluated, please schedule a time to meet with me and discuss; I want you to know that your voice will be heard.

Participation: Your participation is essential. There are multiple ways to participate, but some of these ways include joining class on time, asking questions in class, and sharing solutions when prompted.

In-class reading days: The university has scheduled in-class reading days for February 23/24 and April 7/8. There will be no assessments due and no exams administered on these dates.

Respect: It goes without saying, but I fully expect that we will all be respectful of one another.

Assignments and grades: The grading scheme is designed to be flexible in light of this year's circumstances, with the goal of reducing the excess stress burden.

Problem sets: 50%
There will be a collection of required problems distributed over the course of the semester over seven due dates (namely, every other Thursday). Each problem will be graded holistically on a scale from 0 to 10 (with a score of 10 signifying an immaculate solution with virtually no flaws). Do not expect to receive a perfect score of 10/10 on every problem; scores of 8 and 9 are fabulous, and end-of-semester course grades will reflect this. You are expected to attempt all required problems with good faith effort. Among the problems you complete over the semester with good faith effort, your three lowest-scoring problems will be dropped—the remaining scores will be averaged to compose your problem set percentage. Note that the vast majority of required problems involve the analysis of a real data set. I cannot overstate the importance of reading the documentation associated with those data sets—these documents contain essential information about the problems you're being asked to solve. Collaboration is encouraged, but your assignment should be in your own words (see policy on previous page regarding collaboration).

Exam 1: 25%
A take-home, open-notes, individual-effort examination will be administered on 3/25 covering material from Units 1-3 (due 3/26), comprising several multi-part problems of varying length and difficulty.

Exam 2: 25%
A take-home, open-notes, individual-effort examination will be administered on 5/04 covering material from Units 4-7 (due 5/05), comprising several multi-part problems of varying length and difficulty. Please keep in mind that while material from Units 4-7 will be heavily emphasized, the course content tends to be largely cumulative in nature.

Optional problems: Problem sets and exams will typically feature optional problems. Some credit can be earned for attempting optional problems, and more can be earned for correctness. Such credit is factored in once the course percentages have been aggregated at the end of the semester. To put it another way, it is not possible for completion or non-completion of optional problems to count against you. Coding, matrix algebra, calculus, and simulation studies are fair game for optional problems.

Accommodation: If you have established accommodations with Disability Services, please communicate them to me at your earliest convenience so we can discuss your needs in this course. The contact information for Disability Services is located on the following page: (`https://www.vanderbilt.edu/eeo/disability_services/contact_us.php`).

Tentative topic outline: This is a proposed outline of topics by date. I will speed up and slow down as necessary, so I reserve the right to change this as the course progresses. It will be my responsibility to alert the class of changes.

Unit	Date	Topic
1: Introduction and review	Tu - 1/26	Review
	Th - 1/28	Review and introduction to Stata
2: Continuous outcomes	Tu - 2/02	Simple linear regression (SLR): Introduction
	Th - 2/04	SLR: Further topics
	Tu - 2/09	Multiple linear regression (MLR): Introduction
	Th - 2/11	MLR: Confounding and precision
	Tu - 2/16	MLR: Categorical predictors and effect modification
	Th - 2/18	MLR: Prediction and nonlinearity
	Tu - 2/23	MLR: Weighted least squares
3: Discrete outcomes	Th - 2/25	Binary outcome regression (BOR): Introduction
	Tu - 3/02	BOR: Further topics
	Th - 3/04	Regression of categorical and count outcomes
4: Survival outcomes	Tu - 3/09	Time-to-event outcomes and censoring
	Th - 3/11	Proportional hazards regression (PHR): Introduction
	Tu - 3/16	PHR: Time-varying treatment and competing risks
5: Flexible regression	Th - 3/18	Prediction and cross-validation
	Tu - 3/23	Penalized regression
	Th - 3/25	Regression and smoothing splines
6: Correlated data	Tu - 3/30	Correlated data: Introduction
	Th - 4/01	Longitudinal data: Marginal methods
	Tu - 4/06	Longitudinal data: Conditional methods
7: Additional topics	Th - 4/08	Bayesian methods
	Tu - 4/13	Strategies for missing data
	Th - 4/15	Causal inference: Framework overview
	Tu - 4/20	Causal inference: Methodology (I)
	Th - 4/22	Causal inference: Methodology (II)
	Tu - 4/27	The bootstrap
	Th - 4/29	Sample size and power; wrap-up

Tentative lab outline: This is a proposed outline of lab topics by date.

Week	Date	Topic
Week 1	M - 1/25	Welcome!
Week 2	M - 2/01	Two-sample testing
Week 3	M - 2/08	Simple linear regression
Week 4	M - 2/15	Multiple linear regression
Week 5	M - 2/22	Testing with interaction terms
Week 6	M - 3/01	Binary outcome regression
Week 7	M - 3/08	Multinomial and ordinal regression
Week 8	M - 3/15	Survival analysis
Week 9	M - 3/22	Review for Exam 1
Week 10	M - 3/29	Elastic net
Week 11	M - 4/05	Correlated data
Week 12	M - 4/12	Bayesian regression
Week 13	M - 4/19	Multiple imputation
Week 14	M - 4/26	Review for Exam 2

Problem set due dates: This is a proposed schedule of due dates for each problem. I will alert you of any changes, should they occur.

Set	Date	Required problems	Optional simulation problem
1	Th - 2/04	1, 2	3
2	Th - 2/18	4, 5, 6, 7	8
3	Th - 3/04	9, 10, 11	12
4	Th - 3/18	13, 14, 15, 16	17
5	Th - 4/01	18, 19, 20	21
6	Th - 4/15	22, 23	24
7	Th - 4/29	25	26