BIOS 7345
Advanced Regression Analysis I
Fall 2023

(Version: 09/28/2023)

Instructor:	Andrew Spieker (andrew.spieker@vumc.org) Office: 2525 WEA, 11125
Instructor office hour:	Tuesday 12:00 PM - 1:00 PM (or by appointment)
Class sessions:	Tuesday 10:30 AM - 12:00 PM [2525 WEA, Rm. 11139] Thursday 10:30 AM - 12:00 PM [2525 WEA, Rm. 11139]
Lab section:	Friday, 11:15 AM - 12:15 PM [2525 WEA, Rm. 11139]
TA:	Max Rohde (maximilian.d.rohde@vanderbilt.edu)
TA office hour:	Wednesday 3:00 PM - 4:00 PM [2525 WEA, Rm. 11139]
Course website:	`https://andrewspieker.com/b7345`

Course description: After a review of the most salient topics in linear algebra and multivariate normal theory, we will cover ordinary least squares (full rank and non-full rank cases), weighted least squares, the Gauss-Markov theorem, hypothesis testing, and analysis of variance (ANOVA). In the latter half of the semester, we will cover generalized linear models (e.g., binomial and Poisson models), with focus on estimating equations, algorithms, quasi-likelihood, and sandwich-based variance methods. As time permits, we will cover advanced topics such as cumulative probability models, receiver operating characteristic regression, and conditional logistic regression. This is a methods course; we will use theory as a tool to understand and customize regression methodology and we will use application to illustrate it.

Pre/co-requisites: Exposure to material covered in BIOS 6312 and BIOS 6342 is presumed. BIOST 7345L, features essential course content and must be taken concurrently. R is required.

Required notes: The lecture slides are posted on the course website and are the major source of required material for this course. If you believe there to be an error, please alert me.

Textbook: The texts below are solid reference material but are not "required." Where the texts disagree with lecture material, lecture material takes precedence.

- Seber and Lee. Linear Regression Analysis, 2nd edition. Wiley Series in Probability and Statistics, 2003.
- Wakefield. Bayesian and Frequentist Regression Methods. Springer Series in Statistics, 2013.

Expectations and policies:

Electronic devices: If you choose to use a computer or tablet during class time, I expect you to do so only for the purpose of taking notes.

Notes: I may not cover everything in the notes in extraordinary detail, but I will tell you if there are slides I'd like you to read through outside of class.

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.

Generative AI: Use of generative AI is not permissible; output is presented very confidently even when incorrect and generative AI is often incorrigible in its current state.

Late work: While I expect you to make every reasonable effort to turn work in on time, I understand that life does happen. Late work must be approved by me, in advance when possible.

Accommodations: 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 (`https://www.vanderbilt.edu/eeo/disability_services/contact_us.php`).

Assignments and grades: The grading scheme for the course is shown below.

Homework problems (30%): Each problem attempted with good faith effort will be scored from 6/10 (marginal) to 10/10 (fantastic). Some problems will be optional for students on the MS track. For students on the MS track, the lowest-scoring three problems attempted with good faith effort will be dropped (for PhD-track students, it will be four). Problems not completed with good faith effort will receive a score of 0/10 and will not be dropped. Collaboration is encouraged, but what you turn in *must* be your own (do not turn in anything you do not understand or could not defend). There is value to making honest attempts before attempting to looking up answers in a book or online. Far superior strategies include collaborating with classmates, attending office hours, asking questions in class, etc.

Take-home midterm (30%): Covers material through regularization (open notes; individual-effort). Administered Monday, November 6 (9:00a) and due on Friday, November 10 (5:00p).

Take-home final (30%): Covers material after regularization (open notes; individual effort). Administered Monday, December 11 (9:00a) and due Wednesday, December 15 (5:00p).

Engagement (10%): There are many ways to engage with the course, though non-engagement is in some sense easier to describe (many unexcused absences, not taking notes, staring at your computer screen, making repeated use of your phone during class/lab).

Tentative outline: Below is a tentative outline of course topics, but I will speed up or slow down as necessary. Friday's lab sessions will generally be used to go over problems, review/expand upon ideas introduced in lecture, or catch up on lecture material if need be.

Unit	Date	Topic
1: Foundations	R – 08/24	Introduction and foundations
	F – 08/25	Foundations (continued)
	T – 08/29	Random vectors and matrices (pre-recorded)
	R – 08/31	The multivariate normal distribution
	F – 09/01	Lab
2: Linear models	T – 09/05	Ordinary least squares (full-rank case)
	R – 09/07	Ordinary least squares (full-rank case)
	F – 09/08	Lab
	T – 09/12	Ordinary least squares (rank-deficient case)
	R – 09/14	Ordinary least squares (rank-deficient case)
	F – 09/15	Lab
	T – 09/19	Weighted least squares
	R – 09/21	Hypothesis testing
	F – 09/22	Lab
	T – 09/26	ANOVA and the coefficient of determination
	R – 09/28	Model misspecification
	F – 09/29	No Lab (Vanderbilt Biostatistics Symposium)
	T – 10/03	Confidence regions and prediction
	R – 10/05	Diagnostics
	F – 10/06	Lab
	T – 10/10	Regularization
3: Generalized linear models	R – 10/12	Exponential families
	F – 10/13	Lab
	T – 10/17	Generalized linear models
	T – 10/24	Algorithms and variance estimation
	R – 10/26	Sandwich methods
	F – 10/27	Lab
	T – 10/31	Bootstrap methods
	R – 11/02	Overdispersion and quasilikelihood
	F – 11/03	Lab
	T – 11/07	Hypothesis testing for GLMs
	R – 11/09	Diagnostics for GLMs
	F – 11/10	Lab
	T – 11/14	Additional considerations for binary outcomes
3: Miscellaneous/fun	R – 11/16	Conditional logistic regression
	F – 11/17	Lab
	T – 11/28	Receiver operating characteristic regression
	R – 11/30	Categorical and ordinal outcomes
	F – 12/01	Lab
	T – 12/05	Cumulative probability models (Max)
	R – 12/07	Course review

Problem sets and their due dates: This is a proposed schedule of due dates for collections of problems. Note that a small number of problems are designated as "MS-optional." Students on the MS-terminal track are not required to complete these problems, but may do so for a small amount of extra credit. I reserve the right to alter this schedule as needed, and will alert you of changes should they occur.

Set	**Date**	**Assigned**	**MS-optional**
1	F - 09/01	1-6	6
2	R - 09/07	7-15	14
3	R - 09/14	16-19	–
4	R - 09/21	20-21	–
5	R - 09/28	22-25	–
6	R - 10/05	26,27,29,30	27
7	R - 10/12	31-33	32
8	R - 10/26	34-38	35
9	R - 11/02	39-42	–
10	R - 11/16	43-46	43
11	R - 11/30	47-49	–
12	R - 12/07	50-52	–