Last modified: 17 January 2025

Semester:	**Spring 2025**
Course Number:	**MAR 536**
Course Title:	**Biological Statistics II**
	Instructor:
	Gavin Fay, Associate Professor
	School for Marine Science & Technology
	SMAST East 228; (508) 910-6363; gfay@umassd.edu
Class Location:	SMAST East Room 102
	836 S Rodney French Blvd, New Bedford, MA 02744
	and synchronous via Zoom
Class Time:	Tuesday/Thursday 10:00-11:00
	Wednesday 10:00-11:30
Course website:	gavinfay.github.io/mar536-biolstats2-s25
Student Hours:	R open office: Mondays 1-3 (SMAST-E 235)
	GF: By appointment, Tue 4-5, Thu 1-3

World events, including COVID-19, and other shared and non-visible topics, continue to create uncertainty. These and other topics (both shared and non visible) will at times create stress within our learning community. I want you to succeed in this course and will do what I can to help support your being able to meet our learning objectives. Please communicate with me if there are additional things I can do to help support you.

Code of Conduct: This class follows the Fay Lab Code of Conduct, which applies to all class spaces, including classrooms, Zoom calls, class GoogleDocs or other materials, and student hours. You can also read my lab's lab culture and philosophy, which underpins my approach to teaching and collaboration.

Course Description: Statistical analysis of ecological data. This course provides guided learning in advanced statistical analysis, as applied to ecological research and other fields of marine science. Advanced concepts in probability, statistical model building, hypothesis testing, inference, estimation and prediction, including analyses of example data sets and problems. Students will be required to learn and present topics to the class that are relevant to their graduate research and complete a class project. Course materials will be presented using the statistical programming language R; a computer lab focused on introductory and intermediate R programming will be held for the first half of the semester.

Course Objectives:

1. Self-learning in advanced quantitative concepts and methods
2. Familiarity with advanced statistical methods
3. Experience communicating advanced quantitative topics
4. Experience in statistical programming

Credits: 4

Prerequisites: Students should have taken an introductory graduate statistics course (e.g. BIO 530, MAR 535 Biological Statistics I or equivalent), or seek permission from the instructor.

Required Hardware: Class will take place in a SMAST-East classroom; individual laptop computers will be required to complete in-class and in-lab exercises and coursework. Pair programming approaches are encouraged during class sessions. Please contact the instructor if the computing requirement creates difficulty for your participation in the course; SMAST laptops may be able to be made available for in-class use. The university library also has laptops available for students to borrow.

Required Software: R (free download at http://r-project.org, students should also install Rstudio, an integrated development environment for R, free download at https://posit.co/download/rstudio-desktop/). I will walk through installation and use of this software during our first lab session. It is recommended to update your version of both R and RStudio for the course. Web browser access to R sessions will also be provided through a cloud-based instance of RStudio via posit.cloud.

Course Materials: Materials will be distributed through a dedicated course website and RStudio Cloud project. All lectures and labs will be recorded using Zoom.

Remote participation: To facilitate off-campus attendance and engagement in the course, as well as ensuring course is available and accessible, class sessions will be broadcast via Zoom in addition to in-person meetings. Zoom information can be found in course emails and on the myCourses site. MyCourses will have the links to all class recordings and audio transcripts.

Principal Text:

Zuur, A.F., Ieno, E.N. and Smith, G.M. (2007). Analysing Ecological Data. Springer. 700p. Series: Statistics for Biology and Health. (available free ebook available through the UMass Amherst library system, or hard copy ~$100 online).
Support website for book (https://www.highstat.com/index.php/books2)

Supplementary Texts (others as needed):

Bolker, B.M. 2008. Ecological Models and Data in R. Princeton University Press. (http://ms.mcmaster.ca/~bolker/emdbook/index.html, http://emdbolker.wikidot.com/)

Ismay, C. and Kim, A.Y., 2019. Statistical Inference via Data Science: A ModernDive into R and the Tidyverse. CRC Press.

James, G., Wittem, D., Hastie, T., and Tibshirani, R. (2023). An Introduction to Statistical Learning, with Applications in R. 2nd edition. Springer. (ebook freely available online at https://www.statlearning.com/)

McElreath, R., 2018. Statistical rethinking: A Bayesian course with examples in R and Stan. Chapman and Hall/CRC. (https://xcelab.net/rm/)

Wickham, H., Cetinkaya-Rundel, M., and Grolemund, G., 2023. R for data science. O'Reilly. (https://r4ds.hadley.nz/)

Last modified: 17 January 2025

Course outline and schedule of lectures/labs: (subject to change)

Type	Day	Date	Reading	Topic	Instructor
Lecture	Tue.	Jan 21	McElreath Chap 1.1-1.2	Introduction, statistical rethinking	GF
Lab	Wed.	Jan 22		R Lab 1, Intro to R, working with data	GF
Lecture	Thu.	Jan 23	Bolker 2008 Chap. 4	Probability review	GF
Lecture	Tue.	Jan 28	Zuur et al. Chap. 1-4	data exploration, checking	GF
Lab	Wed.	Jan 29		R Lab 2, Visualizing data	GF
Lecture	Thu.	Jan 30	Zuur et al. Chapter 5	Linear regression review	GF
Lab	Tue.	Feb 4	Hilborn & Mangel Chap 7	Likelihood	GF
Lecture	Wed.	Feb 05		R Lab 3, probability & linear modeling	GF
Lecture	Thu.	Feb 06	Zuur et al. Section 6.1	Extending the linear model (GLM) (Poisson)	GF
Lecture	Tue..	Feb 11	Zuur et al. Section 6.2	GLM 2 (logistic regression)	GF
Lab	Wed.	Feb 12		R Lab 4, data wrangling & model summaries	GF
Lecture	Thu.	Feb 13	James et al. Chapter 5	Resampling methods, Cross-Validation	GF
	Tue.	Feb 18		No class - Monday Schedule	
Lab	Wed.	Feb 19		R Lab 5, Simulation	GF
Lecture	Thu.	Feb 20	James et al. Chapter 7	Nonlinear models, splines	GF
Lecture	Tue.	Feb 25	Zuur et al. Chapter 7	GAMs	GF
Lab	Wed.	Feb 26		R Lab 6, Iteration	GF
Lecture	Thu.	Feb 27		Project updates data explorations	GF
Lecture	Tue.	Mar 4	Zuur et al. Chapter 12	Matrix Algebra Review Midterm distributed	GF
Lab	Wed.	Mar 05		R Lab 7, Functions	GF
Lecture	Thu	Mar 06	Zuur et al. Chapter 12	Principal Components Analysis	GF
				SPRING BREAK	
Lecture	Tue.	Mar 18	Zuur et al. Chapter 8	Linear mixed effects models	GF
Lab	Wed.	Mar 19		R Lab 8, automated reporting & Quarto	GF
Lecture	Thu.	Mar 20		More mixed effects Midterm exam due	GF
Lecture	Tue.	Mar 25		Zero-inflated models	GF
	Wed.	Mar 26		UMass Intercampus Marine Science Symposium	
Lecture	Thu.	Mar 27		Spatial GLMMs	GF
Lecture	Tue.	Apr 1		Project updates	GF
Lab	Wed.	Apr 2		R Lab 9 (working with spatial data)	GF
Lecture	Thu.	Apr 3		GAMs application - guest lecture	
Lecture	Tue.	Apr 08	Zuur et al. Chapter 14	Classification / Linear Discriminant Analysis	GF
Lab	Wed.	Apr 09		Project consult	GF
Lecture	Thu.	Apr 10	Zuur et al. Chapter 9	Trees	GF
Lecture	Tue.	Apr 15	Zuur et al. Chap. 11-15	Ordination methods	GF
Lab	Wed.	Apr 16		Project consult	GF
Lecture	Thu.	Apr 17	James et al. Section 10.3	Cluster Analysis	GF
Presentations	Tue.	Apr 22		Presentations:	student
	Wed.	Apr 23		No class - Monday Schedule	
Presentations	Thu.	Apr 24		Presentations:	student
Presentations	Tue.	Apr 29		Presentations:	student
Lecture	Wed.	Apr 30		Course Review	GF
	Fri.	May 02		**Project report due**	

Evaluation procedures:

1. **An advanced statistical analysis of data** relevant to the student's graduate research topic, including a detailed report and oral presentation (60% of course grade).
   a. 1-2 sentence **project topic idea due Jan 30** (1% of course grade).
   b. Brief (1-2 page) **description of the topic** including a statement of the problem, proposed statistical method, and supplemental reference(s) is **due Feb 13** (4% of course grade).
   c. **Data exploration report** containing data summaries and visualizations of the dataset to be used in the project, including observations of implications of data structure for the analyses to be conducted, and **brief in-class update**. **Due Feb 27**. (10% of course grade).
   d. **In-class verbal update (3 minute presentation)** on project progress. **Apr 1**. (5% of course grade).
   e. **Student project presentation with chapter assignment or supplemental reading on the statistical method used in the project. A 10-minute lecture on the method, and demonstration of the method using the project dataset.** Presentations will be graded based on the instructors' evaluation of accuracy, clarity, and comprehensiveness. Feedback from other students will also be considered in evaluation of presentations (20% of course grade).
   f. Project report (**due May 02**). A written description of the methods and results of the project in the form of a draft scientific manuscript (Introduction, Methods, Results, Discussion). (20% of course grade).
2. **Mid-term examination on statistical concepts and literacy** (20% of course grade). **Take home exam. Exam distributed Mar 4. Due Mar 20.**
3. **Computer lab exercises** (20% of course grade). Completion of short laboratory exercises using R. Lab exercises are due before class on Tuesday the following week. Individual weekly exercises will not be assigned numeric grades, assessment of the R lab component will take the form of **i) completion of the exercises** and **ii) two self-reflection statements** in the middle and at the end of the course asking students to assess their learning in R and ability to use these tools to address data-related and scientific questions. Additionally, the instructor will provide general and specific feedback on R lab components and coding.
4. Extra credit: An additional 2% towards the total course grade can be obtained by students contributing to the #TidyTuesday, a public weekly data visualization project from the Data Science Learning Community (https://github.com/rfordatascience/tidytuesday). To obtain course credit, students should send Gavin Fay an email containing a visualization of analysis of one Tidy Tuesday data set, and a link to a repository containing code that reproduces the analysis and visualization (sharing of work with the #RStats community via Slack or other social media is encouraged but not required for course credit).

Class Policies:

1. Grades will be reduced for all unexcused late submissions of assignments, with a 10% reduction for each day after the deadline. Requests for extended deadlines will be considered up to one day before the deadline. Unexcused assignments submitted later than three days after the deadline will not be graded.
2. Failure to complete any of the requirements for evaluation will result in a score of zero for missing components. At the student's request, and no more than 48 hours after the final exam or class, an incomplete grade may be given only in exceptional circumstances at the discretion of the instructor. The student must be passing the course at the time of the request or be

sufficiently close to passing that the instructor believes that upon completion of the work, the student will pass the course. If the work is not completed within one year of recording the I, the grade will become an F(I).

3. This 4-credit class involves four hours per week of in-class work and an average of eight hours per week of out-of-class work is expected.
4. University policy on academic dishonesty, including plagiarism, applies (see: https://www.umassd.edu/studentaffairs/studenthandbook/academic-regulations-and-procedures/).
5. Assignment submissions should be made via a myCourses website. Students are responsible for all official correspondence sent to their umassd.edu e-mail address.
6. This course, including student participation, will be recorded on video by the instructor and will be available to students in the course for viewing remotely and after each session. As the Zoom recordings will contain student work, please do not share videos beyond the course attendees without the explicit permission of the instructor. Other course materials will be publicly available online.
7. If you have read this far, please use google search to find a picture or gif of an animal that best reflects your opinion of statistics, and send it to Gavin Fay attached to an email with the subject line "Here is a statistical opinion", worth an extra 5 points on the midterm.
8. Available academic support services are available at: www.umassd.edu/nfi/teaching-and-advising/course-syllabus/sample-disability-statement
9. SMAST Code of Conduct and Diversity statement: https://www.umassd.edu/media/umassdartmouth/smast/lab-pdf-files/SMAST-Code-of-Conduct-and-Diversity-Statement.pdf

A full description of Academic Policies associated with this and other UMass Dartmouth courses can be found at: https://www.umassd.edu/provost/resourcesforfaculty/syllabus-language/

Title IX statement: The purpose of a university is to disseminate information, as well as to explore a universe of ideas, to encourage diverse perspectives and robust expression, and to foster the development of critical and analytical thinking skills. In many classes, including this one, students and faculty examine and analyze challenging and controversial topics.

If a topic covered in this class triggers post-traumatic stress or other emotional distress, please discuss the matter with the professor or seek out confidential resources available from the Counseling Center, http://www.umassd.edu/counselling/, 508-999-8648 or -8650, or the Victim Advocate in the Center for Women, Gender and Sexuality, http://www.umassd.edu/sexualviolence/, 508-910-4584. In an emergency contact the Department of Public Safety at 508-999-9191 24 hrs./day.

UMass Dartmouth, following national guidance from the Office of Civil Rights, requires that faculty follow UMass Dartmouth policy as a "mandated reporter" of any disclosure of sexual harassment, abuse, and/or violence shared with the faculty member in person and/or via email. These disclosures include but are not limited to reports of sexual assault, relational abuse, relational/domestic violence, and stalking. While faculty are often able to help students locate

appropriate channels of assistance on campus, disclosure by the student to the faculty member requires that the faculty member inform the University's Title IX Coordinator in the Office of Diversity, Equity and Inclusion at 508-999-8008 to help ensure that the student's safety and welfare is being addressed, even if the student requests that the disclosure not be shared.
For confidential counseling support and assistance, please go to
http://www.umassd.edu/sexualviolence/

Technology Expectations:
I expect that everyone will maintain a classroom conducive to learning. Thus, everyone is expected to behave with basic politeness, civility, and respect for others. In particular, talking in class is okay if it's part of a class discussion, actively troubleshooting code during lab exercises, or with me. Private communications are not permitted. Neither are reading extraneous materials, using electronic equipment off task, or sleeping. Technology is allowed to aid in learning and understanding material. However, please do not use a personal device for any purpose unrelated to our class. All devices should be silenced. Cell phones should be put away when they are not being used for class activities (e.g. taking pictures of work, recording eachother giving practice talks, etc.). If you need to take a call or text during class please step outside. Suggestions for improvement on the class technology policy are welcome at any time. Any concern about the course should be brought first to my attention.

AI, LLMs, and BS (credit to Andrew Heiss)
AI tools are fast changing the landscape when it comes to work. This is especially true for coding, many software are embedding AI into them to assist users produce work and learn. While there is a component of the class related to scientific programming, we are doing this to support our statistical learning – the key objectives in the course are learning about and how to apply statistical methods, not just the computer code to implement them.
I *highly recommend* not using AI tools such as large language models (LLMs) for writing in this class. I am not opposed to LLMs in many situations. For myself, I am beginning to use GitHub Copilot for computer programming-related tasks, most often to assist with tasks *I already know how to do*, or are similar problems to *something I have done before*. Chances are if I write this next year my use cases will have increased. What we do know is that using LLMs requires careful skill and attention and practice, and they tend to be useful only in specific limited cases. Google Docs and Microsoft Word now have built-in text-generation tools where you can start writing a sentence and let the machine take over the rest. ChatGPT and other services let you generate multi-paragraph essays with plausible-looking text. **Please do not use these**. There's a reason most university classes require some sort of writing, like reading reflections, essay questions, and research papers. The process of writing is actually the process of thinking:

> Writing is hard because the process of getting something onto the page helps us figure out what we think—about a topic, a problem or an idea. If we turn to AI to do the writing, we're not going to be doing the thinking either. That may not matter if you're writing an email to set up a meeting, but it will matter if you're writing a business plan, a policy statement or a court case. (Rosenzweig 2023)

Using LLMs and AI to generate this writing will not help you think through the materials. You can create text and meet the suggested word count and finish the assignment, but the text will be meaningless. There's an official philosophical term for this kind of writing: bullshit (Hicks,

Humphries, and Slater 2024; Frankfurt 2005). Philosophical bullshit is “speech or text produced without concern for its truth” (Hicks, Humphries, and Slater 2024). Bullshit isn’t truth, but it’s also not lies (i.e. the opposite of truth). LLMs and AI systems like ChatGPT, Gemini, Claude, and so on are bullshit machines. That might sound hyperbolic, but at a technological level, that’s literally what they do. Do not replace the important work of writing with AI bullshit. Remember that the point of writing is to help crystallize your thinking. Chugging out words that make it look like you read and understand the content will not help you learn. In your project report, I want to see good engagement with the analyses that you have performed and your research questions. I want to see your thinking process. I want to see your personal insights about your work and its importance. I don’t want to see a bunch of words that look like a human wrote them. **That’s not useful for future you**. That’s not useful for me. That’s a waste of time.

I will not spend time trying to guess if your assignments are AI-generated. If you do turn in AI-produced content, I won’t automatically give you a zero. I’ll grade your work based on its own merits. Remember that text generated by these platforms is philosophical bullshit. Since it has nothing to do with truth, it will not—by definition—earn good grades.