BIOL 4400 / 6490 – Applied Biological Data Analysis
Course Syllabus and Schedule
Kennesaw State University – Fall 2024
Instructor: Nicholas S. Green, Ph.D.

1. Basic information

Section	Location	Day(s)	Time (US Eastern)	Modality
BIOL 6490-01	CL 1007	MWF	9:05 – 9:55	Face to face

2. Communication

2.1. Contact info

- **Email:** ngreen62@kennesaw.edu (best contact method). Please allow **1 business day** for a response.
- **Messaging:** Microsoft Teams (KSU) (second best contact method). If I'm online, feel free to message!
- **Phone:** (470) 578-6546 (third best contact method)
- **Office:** Science Building (SC) 331. We can also meet virtually using Microsoft Teams.
- **Office hours:** By appointment

2.2. Course communication

- **D2L** is used extensively for content delivery (e.g., lecture slides, datasets, etc.).
- Grades or other information covered by the Family Educational Rights and Privacy Act (FERPA) can only be communicated using official KSU channels (email or Teams associated with your KSU account) or face-to-face.

3. Course description

Prerequisites: *Admission to the Masters of Science in Integrative Biology (MSIB) program OR Permission of the instructor.*

This course is a survey of data analysis skills and statistical methods essential for careers in biology. This course takes a holistic approach to the data analysis workflow using the open-source software R, including: data management, exploratory analyses, modeling, and reproducible science practices. Statistical topics covered may include generalized linear models, mixed effects models, nonlinear models, and ordination. Students apply techniques learned in class to real biological datasets as a course project.

4. Course objectives

1. Engage with current biology literature to explain the role of statistics in the biological sciences and the ways in which analytical results are communicated.
2. Engage in research using modern statistical methods to investigate biological questions.
3. Manipulate, summarize, display, and analyze data using the open-source environment and language R.
4. Use probability distributions to model and think about biological phenomena.
5. Translate biological hypotheses and ideas into statistical models and vice versa.
6. Conduct exploratory data analyses in support of scientific investigations, particularly to detect and diagnose common problems with biological datasets.
7. Communicate data and analytical results to audiences who may or may not have statistical backgrounds.

5. Textbooks and other materials

5.1. Recommended texts – general

There are no required texts for this course. The following books are provided as recommendations for references on R or statistics. Much of the course is inspired by Bolker (2008), Zuur et al. (2009a), and James et al. (2013). Comprehensive course notes are provided at this link (accessed 2024-08-20).

Bolker BM. 2008. Ecological models and data in R. Princeton University Press, Princeton, New Jersey, USA. *General statistics textbook and introduction to R. Applications are focused on ecology, but the statistical exposition is relevant to any area of biology.*

Dalgaard P. 2008. Introductory statistics with R. Springer Science+Business Media, New York. *General purpose introduction to statistics and to R. Suitable for an introductory stats course, but also useful as a reference for R users of many levels.*

Illowsky B, Dean S. 2018. Introductory statistics. Rice University, Houston, Texas, USA. (Available free online). *Textbook designed for a college-level introductory statistics course.*

James G, Witten D, Hastie T, Tibshirani R. 2013. An introduction to statistical learning with applications in R. Springer Science+Business Media, New York. (7th edition available free online). A*dvanced general statistics textbook that includes both classical methods and more modern techniques such as machine learning.*

Zuur AF, Ieno EN, Meesters EHWG. 2009a. A beginner's guide to R. Springer Science+Business Media, New York. *As the title suggests, this book is a guide to getting started with R for the absolute beginner.*

Zuur AF, Ieno EN, Smith GM. 2007. Analysing ecological data. Springer Science+Business Media, New York. *An advanced statistics textbook that, while focused on ecological applications, is suitable for any biologist.*

5.2. Recommended texts – specialized

The following textbooks and papers are recommended for students needing exposure to more advanced or specialized topics in biological data analysis.

Bolker BM, Brooks ME, Clark CJ, Geange SW, Poulsen JR, Stevens MHH, White J-SS. 2009. Generalized linear mixed models: A practical guide for ecology and evolution. Trends in Ecology & Evolution 24:127–135. DOI: https://doi.org/10.1016/j.tree.2008.10.008.

Borcard D, Gillet F, Legendre P. 2018. Numerical ecology with R. Springer, New York.

Elith J, Leathwick JR, Hastie T. 2008. A working guide to boosted regression trees. Journal of Animal Ecology 77: 802–813. DOI: https://doi.org/10.1111/j.1365-2656.2008.01390.x.

Kéry M. 2010. Introduction to WinBUGS for ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses. Academic Press, Burlington, Massachusetts, USA.

McCune B, Grace JB, Urban DL. 2002. Analysis of ecological communities. MjM software design, Gleneden Beach, Oregon, USA.

Ritz C, Baty F, Streibig JC, Gerhard D. 2015. Dose-response analysis using R. PLoS ONE 10:e0146021. DOI: https://doi.org/10.1371/journal.pone.0146021.

Zuur AF, Ieno EN, Walker NG, Saveliev AA, Smith GM, et al. 2009b. Mixed effects models and extensions in ecology with R. Springer, New York.

5.3. Technology requirements

Access to a computer or laptop capable of running R version ≥4.3. A 64-bit system is highly recommended. The classroom for this course, CL 1007, is equipped with desktop computers for student use.

6. Evaluation and grading

6.1. General information

- Final letter grades are based on performance on exams, homework assignments, and the course project.
- Final course grades will be based on the traditional 10% grading scheme (≥90% = A, 80 – 89% = B, etc.), with percentages rounded to the nearest integer. Per KSU policy there are no +/- grades.

6.2. Grade calculation

Item	% of grade	Description
Pre-assessment	0	This is an ungraded, mandatory assignment designed to measure students' knowledge and skills at the beginning of the semester. Students must complete the pre-assessment in order to be eligible for passing grade in the course.
Homework	20	Periodic homework assignments (approximately biweekly) are used to assess student comprehension and to provide opportunity for practicing key concepts and procedures. Homework is graded on a completion basis. Collaboration between students is ***allowed and encouraged***.
Midterm exam	15	These exams are used to evaluate student understanding and progress at the middle and end of the semester. Exams are take-home, open book, and open note. Collaboration between students is ***not permitted***
Final exam	15	
Course project	50	The most important assignment of the course is a project where students must apply one or more of the statistical methods learned in class to a research question and dataset of their own choosing. A detailed assignment guide with project requirements and grading rubric will be provided early in the semester.

6.3. Attendance policies

6.3.1. Attendance

- Attendance is not required but is highly encouraged. This means that "excused absences" apply to assignment deadlines rather than routine class meetings.
- If an illness or other issue will prevent you from meeting an assignment or exam deadline, please contact me as soon as possible ahead of time so we can make alternative arrangements.

6.3.2. Missed exams and assignments

- Assignments and exams must be turned in on time to be eligible for full credit. Late assignments and exams will be accepted but with a score penalty: 20% for up to 24 hours late; 40% for up to 48 hours late, and so on. The late penalty is waived if you miss the deadline for an excused reason (e.g., death or serious illness in the family, jury duty, or other situations at the instructor's discretion).

7. Other course policies (instructor-specific)

- **Collegial and respectful behavior towards all people is expected at all times.** This does not mean you cannot express your opinions; it means you must be respectful of other people.

8. Academic honesty statement

Every KSU student is responsible for upholding the provisions of the Student Code of Conduct, as published in the Undergraduate and Graduate Catalogs. Section 5c of the Student Code of Conduct addresses the university's policy on academic honesty, including provisions regarding plagiarism and cheating, unauthorized access to university materials, misrepresentation/falsification of university records or academic work, malicious removal, retention, or destruction of library materials, malicious/intentional misuse of computer facilities and/or services, and misuse of student identification cards. Incidents of alleged academic misconduct will be handled through the established procedures of the Department of Student Conduct and Academic Integrity (SCAI), which includes either an "informal" resolution by a faculty member, resulting in a grade adjustment, or a formal hearing procedure, which may subject a student to the Code of Conduct's minimum one semester suspension requirement.

9. Institutional policies and resources

- KSU, BOR, and Federal policies applicable to this course are found at this link.
- Information on help and resources available to students can be found at this link.

10. Course schedule (tentative)

Week	Monday	Topic	Assignments and Exams
1	Aug 12	Syllabus; Intro to R	Pre-assessment due Aug 15 @ 23:59
2	Aug 19	Advanced R: Data manipulation and management	Project: Topic selection due TBD
3	Aug 26	Basic statistics review	Homework 1 due Sept 1 @ 23:59
4	Sept 2	***Monday: No class*** Exploratory data analysis 1: Data description and summarization	
5	Sept 9	Exploratory data analysis 2: Distributions and transformations	Project: Study plan due Sept 10 @ 23:59
6	Sept 16	Exploratory data analysis 3: Exploring multiple variables	Homework 2 due Sept 16 @ 23:59
7	Sept 23	GLM 1: Log-linear and count models	
8	Sept 30	GLM 2: Proportional and binomial models	Midterm due Oct 2 @ 23:59
9	Oct 7	Nonlinear models	Project: Progress check 1 due Oct 11 @ 23:59 Homework 3 due Oct 8 @23:59
10	Oct 14	Mixed effects models	
11	Oct 21	Modeling review	
12	Oct 28	Multivariate 1: Dissimilarity measures and clustering	Homework 4 due Oct 28 @ 23:59
13	Nov 4	Multivariate 2: Ordination	Project: Progress check 2 due Nov 6 @ 23:59
14	Nov 11	Multivariate 3: Multivariate analyses	
15	Nov 18	Wrap-up: planning and presenting your analysis	Homework 5 due Nov 19 @ 23:59
16	Nov 25	*No class – Thanksgiving Break*	
17	Dec 2	**Monday Dec 2: Last day of class (Class presentations)** **Wednesday Dec 4: Class presentations during Final Exam period (tentative)** Course project final deliverable due: Dec 4 @ 23:59 Final exam due: Dec 4 @ 23:59	