**BIO 4490 / 6490 – Applied Biological Data Analysis**
**Course Syllabus and Schedule**
**Kennesaw State University – Fall 2022**
**Instructor: Nicholas S. Green, Ph.D.**

## 1. Basic information

| Section | Location | Day(s) | Time (US Eastern) | Modality |
|---|---|---|---|---|
| BIOL 4490 | Clendenin 1007 | TR | 12:30 PM – 1:45 PM | Face to face |
| BIOL 6490 | Clendenin 1007 | TR | 12:30 PM – 1:45 PM | Face to face |

## 2. Communication

### 2.1. Contact info

- **Email:** ngreen62@kennesaw.edu (best contact method). Please allow **1 business day** for a response.
- **Phone:** (470) 578-6546 (not a good contact method)
- **Messaging:** Microsoft Teams (KSU), Discord (username: **Nick_Green#8554**), or GroupMe (ngreen62@kennesaw.edu)
- **Office:** Science Building (SB) 331. We can also meet virtually using Microsoft Teams or Discord.
- **Office hours:** Tuesday 9:00 – 11:00 AM; Monday 12:00 – 2:00 PM; other times 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:** *Statistical Methods I (STAT 3120) (3 hours) or Biostatistics (STAT 3125) (3 hours) or equivalent.*
**Cross-leveling:** *This course is cross-leveled as BIOL 4490 and BIOL 6490.*

This course is a survey of data analysis skills and statistical methods that are essential for modern biology. The course takes a holistic approach to the data analysis workflow in biology using the open-source environment R, including data management, exploratory data analysis, data modeling, and reproducible science practices. Statistical topics covered include generalized linear models, mixed effects models, non-linear models, and ordination. Students are required to apply techniques learned in class to real or simulated biological datasets as a course project.

## 4. Course objectives

1. Explain the role of statistics in the biological sciences and the ways in which analytical results are communicated.
2. Manipulate, summarize, display, and analyze data using the open-source environment and language R.
3. Use probability distributions to model and think about biological phenomena. Students should come away from the course able to translate biological hypotheses and ideas into statistical statements and vice versa.
4. Conduct exploratory data analyses in support of scientific investigations, particularly to detect and diagnose common problems with biological datasets.
5. Employ modern statistical methods to answer biological questions.
6. Communicate data and analytical results to audiences who may or may not have statistical backgrounds.

## 5. Textbooks and other materials

### 5.1. Recommended texts

*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 built around Bolker (2008).*

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, 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. Technology requirements

Access to a computer or laptop capable of running R version ≥3.0.1 and RStudio version ≥1.4.1106. A 64-bit system is highly recommended.

## 6. Evaluation and grading

### 6.1. General information

- This semester I am trying a system called *ungrading* to assign course grades. This system is more complicated than what most of us are used to, but the main idea is that **your course grade is determined collaboratively by you and the instructor**. This policy has several objectives:
    - To reduce the anxiety and gamesmanship associated with grading. This frees up time and energy for actually learning course material!
    - To offer you a chance and incentive to practice self-guided learning and self-assessment of learning. This is a critical professional skill in science and many other fields.
    - To make the dynamics of the relationship between instructor and students less adversarial and more collaborative.
- Throughout the semester, and at the end of the semester, you will be asked to honestly evaluate your progress in the course and your achievement of the learning objectives. These evaluations will be your justification for the grade you assign yourself at the end of the semester.
    - A rubric will be provided for each self-assessment. You must complete these rubrics and turn them in on time.
    - If you do not turn a rubric, I will assign a grade based on your exams and assignments in the traditional manner (see **6.2**).
    - I reserve the right to adjust student grades if the difference between your self-assessed grade and my assessment of your grade is >15%.

### 6.2. Grade calculation

#### 6.2.1. Grade composition

| Undergraduate students (BIOL 4490) | | Graduate students (BIOL 6490) | |
|---|---|---|---|
| Item | % of grade | Item | % of grade |
| Homework | 20 | Midterm exam | 20 |
| Midterm exam | 20 | Final exam | 20 |
| Final exam | 20 | Course project | 60 |
| Course project | 40 | | |

#### 6.2.2. Grade calculation

- Final letter grades are assigned based our assessment (yours and mine) of how well you met the course learning objectives (see **6.1**)
- Part of the final exam will be a rubric for assessing your overall performance in the course (because there won't be time for the usual post-exam back-and-forth of assessment).
- 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.

### 6.3. Assignment descriptions

- **Homework:** 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. Homework is made available to all students but is only graded for undergraduates registered for BIOL 4490.
- **Midterm and final exams:** 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.
- **Course project:** 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.
    - If a student does not have a research question or dataset, then they can be provided one by the instructor or the student's advisor.
    - Alternatively, students can work with the instructor and their advisor to develop an alternative project such as a "dry run" or power analysis for their intended thesis project.
    - Formal project milestones are staged throughout the semester to facilitate feedback as work on the project progresses.
    - A detailed assignment guide with project requirements and grading rubric will be provided early in the semester.

### 6.4. Attendance and grading policies

#### 6.4.1. Attendance

- Attendance is not required but is highly encouraged. This means that "excused absences" apply to assignment deadlines rather than routine class meetings.
- Excused absences as referred to in this syllabus include the following reasons:
    - Serious illness of self or a close loved one
    - Death of a close loved one
    - Travel or other commitment related to official KSU business (including another course)
    - Interviews for graduate or professional programs
    - Military service or jury duty
    - Closure of the university
    - Other situations at the discretion of the instructor

- Appropriate documentation of an excused absence must be provided upon request.
- If you can foresee an absence, please contact me as soon as possible ahead of time so we can make arrangements for assignments, exams, or activities you might miss.

**6.4.2. Missed exams and assignments**

- Assignments must be turned in on time to be eligible for full credit. Late assignments 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 (see **6.4.1**).
- If you miss an assignment deadline for an excused reason and provide acceptable written documentation for the absence, the assignment will be due at 5:00 PM on the day you return to class (virtually or in person).

**6.4.3. Exam review**

- The review period for an exam, during which you can request score corrections related to question content, is 7 calendar days from the time that exam grades are posted to D2L. After this period no further changes related to question content will be made.
- Challenges to the content of exam questions must be made during the 7-day review period described above. Any challenge should be accompanied by a written explanation of the issue with the question. The explanation should include appropriate documentation (e.g., a recent textbook or journal article citation). One to three sentences per question is usually enough explanation.
- Score corrections for reasons not related to exam content (e.g., arithmetic error in grading) can be requested at any time up until 5:00 PM on the day before the Final Exam.

**7. Other course policies (instructor-specific)**

- Students must abide by the KSU Student Code of Conduct at all times.
- **Collegial and respectful behavior towards all people is expected.** This does not mean you cannot express your opinions; it means you must be respectful of other people.
- Children may not accompany parents into the classroom without the explicit permission of the instructor. Disruptive children will not be allowed to remain in the classroom.

**8. Institutional policies and resources**

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

## 9. KSU COVID-19 related polices

**9.1. Important reminders**

- Faculty notification of a positive rest result is the responsibility of the individual student as with any other illness. Faculty members may request documentation from students as they would for any other illness-related absence.
- Students who test positive should be accommodated in respect to making up work missed during a quarantine/isolation period in the same manner they would be accommodated for any other illness.
- The method or approach to providing content and makeup work for the students is at the faculty member's discretion.
- Faculty should not conduct contact tracing or notify the class of a positive case.
- Students, faculty, and staff should not be asked about their vaccination status.

**9.2. Isolation/quarantine guidance**

- Visit kennesaw.edu/studenthealth for more information on quarantine and isolation protocols.

**10. Course schedule (tentative)**

| Week | Days | Topic | Assignments and Exams |
|---|---|---|---|
| 1 | Aug 16, 18 | Statistics in biology | |
| 2 | Aug 23, 25 | Introduction to R | Pre-assessment due Friday August 26 @ 11:59 PM |
| 3 | Aug 30, Sept 1 | Data manipulation and management | |
| 4 | Sept 6, 8 | Exploratory data analysis 1: Data description and summarization | Homework 1 due Sunday Sept 11 @ 11:59 PM |
| 5 | Sept 13, 15 | Exploratory data analysis 2: Distributions and transformations | Project: Topic selection due Monday September 12 @ 11:59 PM |
| 6 | Sept 20, 22 | Exploratory data analysis 3: Exploring multiple variables | Homework 2 due Sunday Oct 2 @ 11:59 PM |
| 7 | Sept 27, 29 | Statistical problems<br>Planning your analysis | |
| 8 | Oct 4, 6 | GLM 1: Introduction; log-linear models<br>GLM 2: Models for counts | Project: study plan due Friday Oct 7 @ 11:59 PM |
| 9 | Oct 11, 13 | GLM 3: Models for binary outcomes and proportions | Midterm due Monday Oct 10 @ 11:59 PM |
| 10 | Oct 18, 20 | GLM 4: Wrap up and extensions<br>Nonlinear models | Homework 3 due Sunday Oct 23 @ 11:59 PM |
| 11 | Oct 25, 27 | Mixed effects models | Project: Progress check 1 due Sunday Oct 30 @ 11:59 PM |
| 12 | Nov 1, 3 | Multivariate 1: Dissimiliarity | |
| 13 | Nov 8, 10 | Multivariate 2: Ordination | Homework 4 due Sunday Nov 13 @ 11:59 PM |
| 14 | Nov 15, 17 | Multivariate 3: | Project: Progress check 2 due Sunday Nov 20 @ 11:59 PM |
| 15 | Nov 22, 24 | *No class – Thanksgiving Break* | |
| 16 | Nov 29, Dec 1 | Course project presentations | |
| 17 | | Course project final deliverable due: Sunday Dec 4 11:59 PM<br>Course project presentations (if needed): Thursday Dec 8 1:00 – 3:00 PM<br>Final exam due: Thursday Dec 8: 11:59 PM | |
| 18 | | Final grades due to Registrar: Thursday Dec 15 12:00 PM | |