Biology 204
Biological Data Analysis
Fall 2016

Meeting times and locations

Lectures: Mon, Wed 1:25-2:40pm, Fri 1:25PM - 2:15PM (recitation), Gross 103

Instructors

Dr. Paul M. Magwene, Office: 4103 FFSC, Phone: 613-8159, Email: paul.magwene@duke.edu

Teaching assistant

Cullen Roth, Office: 3135C FFSC, Phone: 668-3755, Email: cullen.roth@duke.edu

Synopsis

Principles and applications of statistics and scientific computing in biology, with emphasis on genetics, molecular biology, ecology and environmental science. Topics include: the presentation of biological data, summary statistics, probabilities and commonly-applied probability distributions, the central limit theorem, statistical hypothesis tests, errors and power, tests using the z- and t-distributions, analysis of variance, correlation and regression, non-parametric tests.

Prerequisites

No prerequisites are enforced at registration but students are assumed to be familiar with the material covered in the mathematical courses required for the Duke Biology BS degree (see Duke Biology Undergraduate Requirements).

Course website

See the Biology 204 Sakai Website.

Required Texts

[1] Diez, D. M., C. Barr, and M. Çetinkaya-Rundel. 2015. OpenIntro Statistics: Third Edition. Free PDF available at the OpenIntro website or as an inexpensive paperback from Amazon.com.

Office hours

The course TA will set up regular office hours.

The instructor is available *by appointment*. Please email, call, or meet with the instructor after class to setup an appointment.

Grading

We will have approximately 9-12 quizzes and approximately 9-12 problem sets (exact numbers of each will depend on the pace we proceed through the materials). Grades will be determined based on the sum of the total quiz and problem set scores. An appropriate grading curve will be applied when determining final grades.

Format

The course is based around lectures, in-class computational exercises and simulations, and problem sets. Lectures will introduce key statistical, mathematical, and computational concepts. In-class computational exercises are meant to illustrate those concepts with applications to real or synthetic data and provide concrete instructions on how to implement these concepts in a Python based scientific computing environment.

Learning objectives

1. Know how to explore, summarize, describe and display your data
2. Know what a random variable is
3. Know what conditional, marginal and joint probabilities are
4. Know what the Normal, t-, F, χ^2, Poisson and Binomial distributions are and how they are used
5. Know what the Central Limit Theorem states
6. Know the rationale behind hypothesis testing
7. Know what Type I and Type II errors are, what a p-value is, and what the "power" of a test refers to
8. Know what correlation and regression are and how they differ
9. Know what analysis of variance (ANOVA) is and how it works
10. Know when to perform a t-test, an ANOVA or a regression for a given problem
11. Know how to perform goodness-of-fit tests using χ^2
12. Know about random sampling, replication, and pseudoreplication

Expectations and policies

Students are expected to:

1. **Come to class on time.**
2. Read assigned readings *before* lecture.
3. Submit homework and projects by their due dates.
4. Make sure you understand what's going on with every lecture. Lectures and related assignment, build on earlier material so failure to grasp key concepts from a single lecture can easily send you off the rails in later lectures. If something is unclear please discuss it with the instructor or TA!
5. Adhere to the Duke Community Standard (see "Academic Integrity" below). A zero tolerance policy is in effect: any assignments violating the Standard will receive a zero grade.

Missed class time and late submissions

Students are expected to attend and participate in every class session. In the case of illness or extraordinary personal circumstance students should notify the instructors and their academic deans by means of a Short-Term Illness Notification Form (STINF). In the case of long-term medical (three or more STINFs) or personal circumstances students should request a letter from their academic dean. Note that STINFs only apply to in-class work (i.e. quizzes). "Make-up" quizzes will not be offered (see explanation of quiz grades above).

Students who will miss a class session due to religious observance should submit a Religious Observance Notification Form *no later than one week prior to the date of the holiday.*

Students who are members of a varsity athletic team, and who may be required to miss some class time, should provide the instructors with a copy of their spring semester travel schedule during the first week of classes. Students should also complete an online "Notification of Varsity Athletic Participation (NOVAP)" form *no later than one week prior to each absence.*

Students who are seniors, and will be travelling for interviews, should notify the instructor *at least one week in advance about missed class time.*

Students with excused absences are still expected to complete and submit homework assignments on time. Homework assignments that are submitted late will receive half credit if submitted within 24 hours of the due date, or zero credit thereafter.

Academic Integrity

All students are expected to adhere to, and have an obligation to act, in accordance with the Duke Community Standard. For more information about the community standard please refer to the documents available at: https://studentaffairs.duke.edu/conduct/about-us/duke-community-standard. Strict adherence to the plagiarism policy described in the Community Standard will be observed. Any violations of the community standard will be referred to the undergraduate judicial board. Students are encouraged to study together and discuss the course material.

Topic Syllabus, Fall 2016

Week	*Dates*	*Topic*	*Notes*
1	Aug 29 - Sep 1	Introductory Materials	–
2	Sep 5 - 9	Getting up to speed with R	–
3	Sep 12 - 16	Visualizing and describing distributions I	–
4	Sep 19 - 23	Visualizing and describing distributions II	–
5	Sep 26 - 30	Regression and curve fitting	–
6	Oct 3 - 7	Measures of similarity; clustering	No class: Oct 5
7	Oct 10 - 14	More advanced R programming concepts	No class: Oct 10
8	Oct 17 - 21	Probability I	No class: Oct 19, 21
9	Oct 24 - 28	Probability II	–
10	Oct 31 - Nov 4	Statistical distributions	–
13	Nov 21 - 25	Statistical inference II	No class: Nov 23, 25
14	Nov 28 - Dec 2	Statistical inference III	–
15	Dec 5 - 9	Statistical inference IV	–