03-702: Statistical Computing for Biological Sciences

Fall 2024

**Instructor:**  En Cai

 Office: MI 612

 Email: encai@andrew.cmu.edu

**Class time:** Mondays and Fridays, 11:00-12:20 pm, DH 1212

**TA:** Suzi Kim (suzik@andrew.cmu.edu), Elianna Lai (eliannal@andrew.cmu.edu)

**Course objectives:** In biology, we often work with a diverse range of data types. To interpret these data effectively, it is essential to have a foundational understanding of statistics and the necessary tools for statistical analysis. The goal of this course is to equip biology students with practical statistical analysis skills. It will focus on fundamental statistical concepts, hypothesis testing, and essential tools for data analysis, with a particular emphasis on applying these skills in Python for biological data analysis.

**Coursework:** Students will be graded on three components:

1. Participation (25%) – Students are required to participate in the class which will account for 25% of their final grade.
2. Homework (50%) – Students are required to complete weekly homework and turn in their homework on time. This will account for 50% of their final grade.
3. Final project (25%) -- Students are required to complete a final project at the end of the course. This will account for 25% of their final grade.

Policies

**Absences:** Students are required to attend class and this accounts for 25% of their final grade. One excused absence is provided by default. Additional absences can be excused for extenuating circumstances. Absences for religious observances must be submitted by email to the instructor during the first two weeks of the course.

**Academic integrity:** Study groups can be helpful in understanding course material, but students are responsible for preparing their own assignments independently. All material turned in must be done independently unless explicitly indicated on the assignment handout. The University Policy on Academic Integrity can be found here: <http://www.cmu.edu/policies/student-and-student-life/academic-integrity.html>.

**Homework**: Homework will be assigned every week and will be submitted online. One late homework is allowed (you can submit one assignment up to 7 days late). Additional late submissions will incur the following grade penalties:

Submitted within 48 hours of the deadline - 10% reduction in the homework assignment grade.

Submitted 2-7 days after the deadline - 50% reduction in the homework assignment grade.

Submitted more than 7 days after the deadline will receive a zero grade.

**Course Calendar (subject to revision)**

The planned class sessions are as follows:

Oct 21. Lecture 1: Course overview and introduction to statistical analysis

Oct 25. Lecture 2: Discrete distribution and Normal distribution

Oct 28. Lecture 3: Normal distribution and distribution functions

Nov 1. Lecture 4: Other continuous distribution functions

Nov 4. Lecture 5: Hypothesis Tests Part 1

Nov 8. Lecture 6: Hypothesis Tests Part 2

Nov 11. Lecture 7: Test of Means of numerical data Part 1

Nov 15. Lecture 8: Test of Means of numerical data Part 2

Nov 18. Lecture 9: Test on Categorical Data

Nov 22. Lecture 10: Statistical Modeling Part 1

*Nov 25. No Class – Thanksgiving*

Nov 29. Lecture 11: Statistical Modeling Part 2

Dec 2: Lecture 12: Review and Course wrap-up

Dec 6: Lecture 13: Final Project Due

**Diversity and Inclusion:** At Carnegie Mellon University, students, faculty, staff, and researchers come from diverse backgrounds. Students from diverse backgrounds and perspectives should be able to benefit from this course. In order to maintain our diversity, we must not discriminate against others because of their personal identities. In the event that you suspect there has been a bias based on an individual's identity, you can contact me or the Center for Student Diversity and Inclusion at csdi@andrew.cmu.edu, (412) 268-2150.

**Textbook and Reading:**

1. An Introduction to Statistics with Python, by Thomas Haslwanter.
2. Computing Skills for Biologists, by Stefano Allesina and Madlen Wilmes.