Course syllabus

Applied Statistics for Life Sciences

STAT218

January 2024

Statistics plays a crucial role in the sciences: statistical techniques provide a means of weighing quantitative evidence derived from observation and experimentation in the face of uncertainty. Statistical thinking and data analysis also facilitate discovery, exploration, and hypothesis generation. Likewise, the sciences play a crucial role in statistics: technological and knowledge innovations in methods of scientific investigation motivate the development of new statistical methods for data analysis.

This class aims to provide a hands-on introduction to common statistical methods used almost universally across the sciences — descriptive and graphical techniques, inferential methods for comparing population means, analysis of categorical data and contingency tables, and linear regression — while drawing on examples from the life sciences to help illuminate the potential for application in students' chosen field(s) of study and providing basic training in the use of statistical software. The class also creates a unique opportunity for students to interact broadly and make connections across majors and class standing.

Course information

Instructor: Trevor Ruiz (he/him/his) [email]

Class meetings:

- [Section 05/8273] 2:10pm — 4:00pm TR Construction Innovations Center Room C202
- [Section 06/8274] 4:10pm — 6:00pm TR Business Room 113

Class meetings will comprise a mixture of lecture, lab activities, class activities, and discussion.

Office hours: 12:10pm — 2:00pm MW 25-236 or Zoom [by appointment].

These times are partitioned into 10 minute intervals that you can schedule via the appointment link above; this system is intended to minimize waiting times and guarantee one-on-one

availability. Slots can be scheduled anywhere from 7 calendar days to 10 minutes in advance. While drop-ins are welcome, I can't guarantee availability outside of scheduled times.

Catalog Description

Data collection and experimental design, descriptive statistics, confidence intervals, parametric and non parametric one and two-sample hypothesis tests, analysis of variance, correlation, simple linear regression, chi-square tests. Applications of statistics to the life sciences. Substantial use of statistical software.

Prerequisite: MATH 96; or MATH 115; or appropriate Math Placement Level.

Fulfills GE Area B4 (GE Area B1 for students on the 2019-20 or earlier catalogs); a grade of C- or better is required in one course in this GE area.

Materials

To access course materials, engage in class, and complete assignments, you'll need an internet-connected (a) laptop or (b) tablet with a keyboard — the keyboard is necessary since we will do some web-hosted computation and you will be expected to type assignments. I'll let you know in advance when you need to bring a laptop/tablet to class. Besides your personal computer, all materials are free.

Textbooks:

1. Vu and Harrington (2020). Introdutory Statistics for the Life and Biomedical Sciences, First edition. A PDF and tablet-friendly version are available for free online at the link above. This will be our primary reference and we will cover chapters 1 – 2, 4 – 6, and 8.
2. Van Belle, Fisher, Heagerty, and Lumley (2004). Biostatistics: a methodology for the health sciences. Wiley. A PDF can be obtained through the Kennedy Library via the link above. This text provides a thorough introduction to biostatistics (statistics for life sciences) and is an excellent reference for more depth of coverage. Select readings will be assigned from this book.
3. Douglas *et al.* (2023). An Introduction to R. This online book covers a variety of introductory topics pertaining to R/RStudio: installation, packages, files and directories, objects, functions, data types, data structures, graphics, basic statistics, markdown, and version control. Select readings will be assigned from this book.

Computing: computing will be hosted online via a [posit.cloud workspace]. To access the workspace, you'll need to create a (free) posit.cloud account — open the invitation by clicking the link above and follow prompts. Activities and assignments will be distributed via this workspace. Please be aware that any files you create on the workspace will be visible to admins.

Course notes: notes to supplement readings will be posted as needed on the course website.

Learning outcomes

The course aims to enable you to demonstrate the following abilities.

- [L1] design a data collection scheme based on simple random sampling or simple experimental designs
- [L2] distinguish between observational studies and experiments and understand the limitations (practical and consequential) of each
- [L3] summarize data using graphical and numerical techniques
- [L4] construct and interpret confidence intervals for means and differences between means for independent and paired samples
- [L5] conduct parametric and non-parametric two-sample hypothesis tests for means
- [L6] construct and interpret a confidence interval for a single proportion
- [L7] conduct Chi-square goodness-of-fit tests and tests for independence
- [L8] distinguish between case-control and cohort studies and compute relative-risk and odds in the appropriate settings
- [L9] perform analysis of variance tests and post-hoc comparisons for completely randomized designs
- [L10] use simple linear regression to describe relationships between variables
- [L11] apply one or more methods from the course to your major field of study

In addition, you will learn to perform simple analyses and computations in R and can expect to attain a basic level of proficiency with the software; however, as this is not a programming class, emphasis will be placed on obtaining and interpreting relevant outputs in the context of the analyses indicated above.

Assessments

Attainment of learning outcomes will be measured by performance on homework assignments, tests, and a final oral exam.

Homeworks are your opportunity to practice using course concepts and methods covered in class and comprise both graded and ungraded questions, marked as such. Graded questions will be largely data analytic and will correspond to specific marked learning outcomes; responses will be assessed as satisfactory (S) or needing improvement (NI) according to whether they are fully correct, and qualitative feedback will be offered if an assessment of NI is given. Ungraded questions are entirely for your benefit as extra practice. **You can submit revisions** to any graded responses needing improvement for reassessment.

Tests are your opportunity to demonstrate that you've synthesized course material and achieved learning outcomes. Tests will comprise sets of questions, some conceptual and some data analytic, corresponding to specific marked learning outcomes. Responses will be assessed as satisfactory (S) or needing improvement (NI) according to whether they are fully correct, and qualitative feedback will be offered if an assessment of NI is given. You will be given a 24 hour window to complete each test, and **you can submit revisions** to responses needing improvement.

The **final oral exam** will require you to find an application of course material in the field of your major and present it as a case study in 5 minutes. You will be expected to describe the research question, study design, analysis, and conclusion, and answer 1-2 questions. These will be carried out in private and scheduled during the final exam time. This assignment applies to learning outcome L11 only, and you will receive an assessment of 'fully met', 'partly met', or 'unmet' for that outcome depending on your presentation and ability to answer questions. A rubric will be provided in advance to clearly define the expectations associated with each possible assessment.

Letter grades

Students will receive a score for each learning outcome representing the weighted proportion of questions matched with that outcome that received a satisfactory assessment across all assignments, with relatively more weight given to questions from tests. The outcome will be determined as 'fully met' if the weighted proportion is at least 0.8, 'partly met' if the weighted proportion is between 0.5 and 0.8, and 'unmet' otherwise. To receive a passing grade in the class, at least six outcomes must be partly met. Letter grades are then defined as follows:

Grade	Minimum number of partly + fully met outcomes	Minimum number of fully met outcomes
A	10	9
A-	10	8

Grade	Minimum number of partly + fully met outcomes	Minimum number of fully met outcomes
B+	8	7
B	8	6
B-	8	5
C+	6	5
C	6	4
C-	6	3
D+	6	2
D	6	1
D-	6	0

Please note that these definitions are tentative and potentially subject to change. Please also note that failure to adhere to course policies — particularly collaboration, academic integrity, and attendance policies — may result in a lower letter grade than would otherwise be assigned.

Tentative schedule

Subject to change.

Week	Topics	Readings (V&H)	Assignments
1 (1/8/24)	Introduction to statistical thinking and study design	1.1	HW1
2 (1/15/24)	Data, data types, and data collection	1.2	None
3 (1/22/24)	Foundations for inference	1.4 – 1.5; 4.1	HW2
4 (1/29/24)	Interval estimation and hypothesis tests	4.2 – 4.3	Test 1
5 (2/5/24)	Two-sample inference for numerical data; nonparametric alternatives	5.1 — 5.3	HW3
6 (2/12/24)	Comparing means with analysis of variance	5.4	HW4
7 (2/19/24)	Inference for categorical data: proportions	8.1 — 8.2	Test 2
8 (2/26/24)	Inference for categorical data: chi-square tests	8.3 — 8.4	HW5
9 (3/4/24)	Simple linear regression: model framework and estimation	6.1 — 6.3	HW6

Week	Topics	Readings (V&H)	Assignments
10 (3/11/24)	Simple linear regression: inference	6.4 — 6.5	Test 3
Finals (3/18/24)	N/A	N/A	Oral exam

Course policies

Collaboration

Collaboration within the class and across class sections is allowed and encouraged on homework assignments, subject to certain conditions outlined in the paragraph below. Collaborations should not include individuals outside of the class. Students collaborating with a group are expected to prepare their own assignment submissions, in their own words and writing, and should to indicate their collaborators in writing on their submission. This is limited to peers that were consulted closely in completing the assignment; brief or passing interactions are not considered collaborations.

A collaboration is a shared effort. Students that choose to work together on homework assignments are expected to make material contributions towards producing one or more shared answers or solutions. Material contributions might include participation in discussions, critique of a proposed solution, or presentation of a problem approach. In the absence of such contributions, submitting solutions prepared in a group is not appropriate. The best way to adhere to this policy is to attempt problems individually before consulting others and exchanging work.

Attendance

Regular attendance is essential for success in the course and required per University policy. Each student may miss two class meetings without notice but additional absences should be excusable and students should notify the instructor. Unexcused absences may negatively impact course grades.

Communication and email

Students are encouraged to use face-to-face means of communication (office hours, class meetings, appointments) when possible. Email may be used on a secondary basis or when a written record of communication is needed. Every effort is made to respond to email within 48 weekday hours — so a message sent Thursday or Friday may not receive a reply until Monday or Tuesday. Time-sensitive messages should be identified as such in the subject line. For non-time-sensitive messages, please wait one week before sending a reminder.

Time commitment

STAT218 is a four-credit course, which corresponds to a minimum time commitment of 12 hours per week, including lectures, reading, assignments, and study time. Some variability in workload by week should be expected, and most students will need to budget a few extra hours each week. However, students can expect to be able to meet course expectations with a time commitment of 12-16 hours per week. Considering that lecture accounts for four hours per week, students should anticipate devoting 8-12 hours outside of class.

Assignment scores and final grades

Every effort will be made to provide consistent, fair, and accurate evaluation of student work. Students should notify the instructor of any suspected errors or discrepancies in evaluation promptly on an assignment-by-assignment basis (*i.e.,* not at the end of the quarter) to guarantee consideration. Final (letter) grades will only be changed in the case of clerical errors. Attempting to negotiate scores or final grades is not appropriate.

Per University policy, faculty have final responsibility for grading criteria and grading judgment and have the right to alter student assessment or other parts of the syllabus during the term. If any student feels their grade is unfairly assigned at the end of the course, they have the right to appeal it according to the procedure outlined here.

Deadlines and late work

Each student may turn in two homework assignments up to one week late without penalty at any time during the quarter and without notice. Subsequently, late work will incur a penalty in final grade calculations unless an extension is granted in advance. As a general policy, late work will be not accepted beyond one week after the original due date. Deadlines for tests are strict.

These policies are intended to provide you with some flexibility to work around unforeseen circumstances while maintaining accountability for completing coursework in a timely manner. That said, if any circumstances arise that the policies do not accommodate well, please let me know and I will do my best to work with you to keep you on track in the course. Exceptions may be granted for significant and unforeseen challenges (medical absences, family emergencies, and the like).

Accommodations

It is University policy to provide, on a flexible and individualized basis, reasonable accommodations to students who have disabilities that may affect their ability to participate in course activities or to meet course requirements. Accommodation requests should be made through the Disability Resource Center (DRC).

Conduct and Academic Integrity

Students are expected to be aware of and adhere to University policy regarding academic integrity and conduct. Detailed information on these policies, and potential repercussions of policy violations, can be found via the Office of Student Rights & Responsibilities (OSRR).

Instances of academic dishonesty will be reported to OSRR and will result in penalty ranging from credit reduction to grade reduction to course failure, depending on the severity of the situation.

Copyright and distribution of course materials

All course materials, including handouts, homework assignments, lab assignments, study guides, course notes, exams, and solutions are subject to copyright; students are not permitted to share or distribute any course materials without the written consent of the instructor. This includes, in particular, uploading materials or prepared solutions to online services and sharing materials or prepared solutions with students who may take the course in a future term. Transgressions of this policy compromise the effectiveness of instruction and assessment and do a disservice to current and future students.