Course syllabus

Applied Statistics for Life Sciences

STAT218

April 2024

Statistics plays a crucial role in the sciences: statistical techniques provide a means of weighing quantitative evidence derived from observation and experimentation while accounting for uncertainty. Statistical thinking and data analysis also facilitate discovery, exploration, and hypothesis generation. 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.

### Course information

**Instructor:** Trevor Ruiz (he/him/his) [[email](mailto:truiz01@calpoly.edu)]

**Class meetings:**

* [Section 05] 12:10pm — 2:00pm MW Construction Innovations Center Room C100
* [Section 06] 2:10pm — 4:00pm MW Construction Innovations Center Room C100

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

**Office hours:** 8:10am — 11:00am Mondays 25-236 or Zoom [[by appointment](https://calendly.com/tdruiz/office-hour)].

These times are partitioned into 15 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

You’ll need an internet-connected laptop or tablet (a keyboard is necessary since we will do some web-hosted computation and you will be expected to type assignments). You should expect to bring your laptop or tablet to every class meeting.

**Computing**: use of R/RStudio will be hosted online via a posit.cloud workspace [[link to join](https://posit.cloud/spaces/497811/join?access_code=UXcfuEu2vkl3e2KOA4d4IU3UejqhsSqMMcqbyLT7)]. To access the workspace, you’ll need to create a posit.cloud account and purchase a $5/month student plan.

**Textbook**: Vu and Harrington (2020). [Introdutory Statistics for the Life and Biomedical Sciences](https://www.openintro.org/book/biostat/), 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.

**Course notes**: course notes will be posted as slides on the course website.

**Other references**:

* Van Belle, Fisher, Heagerty, and Lumley (2004). [Biostatistics: a methodology for the health sciences](https://csu-calpoly.primo.exlibrisgroup.com/permalink/01CALS_PSU/1qh1nk7/alma991005991838302920). 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.
* Douglas *et al.* (2023). [An Introduction to R](https://intro2r.com). 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.

### Learning outcomes

This course aims to support you in developing 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

Emphasis is placed on conceptual fluency, application, and interpretation. In addition, you will learn to perform simple statistical analyses in R and can expect to develop a basic familiarity with the software; however, as this is not a programming class, the R environment will not be discussed in any detail and you will only learn to use a handful of commands.

### Assessments

Attainment of learning outcomes will be measured by performance on homework assignments, tests, and a short project with an oral assessment in lieu of a final exam.

* **Homework assignments** will be given at the end of every class meeting and will comprise two practice problems due by the next class meeting. These are your opportunity to practice applying course concepts and methods covered in class and will help you to keep current with the pace and content of the lectures.
* **Tests** will be given every 2-3 weeks and will comprise roughly 10-20 problems each. These are your opportunity to demonstrate that you’ve synthesized course material and achieved learning outcomes, and you will have approximately 48 hours to complete each test. One round of revisions will be allowed for each test in which you can make up full credit for any problems answered incorrectly in your initial attempt.
* A **project** with an oral assessment will be given in place of a final exam. However, **you will need to be available in person during the scheduled final exam time**, as this is when the oral assessment will take place.

Every assessed problem will be matched to one of the learning outcomes L1-L10. All submitted work will be assessed on a question-by-question basis as satisfactory (S) or needing improvement (NI) according to whether responses are fully correct. The percentage of problems matched to a particular learning outcome for which you receive a satisfactory assessment provides a measure of your attainment of that learning outcome. These percentages form a basis for determining your course grade (see below).

Due to limited resources we will only provide qualitative feedback on a small subset of assessed questions, and only when an assessment of NI is made. As such, it is your responsibility to seek the feedback you need to correct your understanding where needed via class engagement, office hours, peer consultation, further study, and [[tutoring resources](https://statistics.calpoly.edu/content/tutoring)].

### Letter grades

Students will receive a score for each learning outcome representing the (possibly weighted) proportion of questions matched with that outcome that received a satisfactory assessment across all assignments. The outcome will be assessed as follows:

* ‘fully met’ if the proportion is at least 0.8
* ‘partly met’ if the proportion is between 0.5 and 0.8
* ‘unmet’ otherwise

You will receive periodic email summaries of your progress on each learning outcome. **To receive a passing grade in the class, at least six outcomes must be either partly or fully met.** Subject to this condition, letter grades are then defined as follows:

| Grade | Number of fully met outcomes |
| --- | --- |
| A | 10 |
| A- | 9 |
| B+ | 8 |
| B | 7 |
| B- | 6 |
| C+ | 5 |
| C | 4 |
| C- | 3 |
| D+ | 2 |
| D | 1 |
| D- | 0 |

Please note that these definitions are tentative and potentially subject to change; however, I will not make the grading requirements more stringent under any circumstances.

Please also note that failure to adhere to course policies may result in a lower letter grade than would otherwise be assigned.

### Tentative schedule

Subject to change.

| Week | Topics | Readings (V&H) | Assessments |
| --- | --- | --- | --- |
| 1 (4/1/24) | Introduction to statistical thinking and study design | 1.1 |  |
| 2 (4/8/24) | Data, data types, and data collection | 1.2 |  |
| 3 (4/15/24) | Descriptive statistics and graphical summaries | 1.4, 1.5, 1.6 | Test 1 [L1, L2, L3] |
| 4 (4/22/24) | Foundations for inference | 4.1, 4.2 |  |
| 5 (4/29/24) | One-sample inference for numerical data | 4.3, 5.1 |  |
| 6 (5/6/24) | Two-sample inference for numerical data | 5.2, 5.3, 5.4 | Test 2 [L4, L5] |
| 7 (5/13/24) | Nonparametric tests; analysis of variance | 5.5 |  |
| 8 (5/20/24) | Post-hoc inference in ANOVA; intro to categorical data analysis | 8.1 | Test 3 [L6, L9] |
| 9 (5/27/24) | Categorical data analysis and contingency tables | 8.3, 8.5.1, 8.5.3 |  |
| 10 (6/3/24) | Simple linear regression | 6.1, 6.2, 6.4, 6.5 | Test 4 [L7, L8, L10] |
| Finals (6/10/24) | N/A | N/A | Oral project assessment [L11] |

### Course policies

#### 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 class meetings account for four hours per week, students should anticipate devoting 8-12 hours outside of class. If you are spending considerably more time than this on a regular basis, please let me know.

#### 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](https://academicprograms.calpoly.edu/content/academicpolicies/class-attendance) and students should notify the instructor. Unexcused absences may negatively impact course grades.

#### Deadlines and extensions

A one-hour grace period is applied to all deadlines. Work submitted more than one hour after a deadline is considered late. Policies regarding late work are as follows:

* You may turn in as many as four homework assignments up to 48 hours late without penalty at any time during the quarter and without notice. Subsequently, late work may incur a penalty in final grade calculations.
* Late submissions are not allowed for tests. You are expected to plan ahead in order to meet test deadlines; I recommend putting the dates in your calendar at the beginning of the quarter.
* Exceptions may be granted for significant and unforeseen challenges (medical absences, family emergencies, and the like).

Extensions may be arranged as needed if warranted by the circumstances and should be requested by email. When requesting an extension, you should explain why it is needed; it is at my discretion to grant the extension or not based on the reason provided. **Extensions must be arranged at least 24 hours in advance of the original deadline**; requests made after this time will not be considered as a general rule.

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.

#### Academic integrity

You 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](https://osrr.calpoly.edu) (OSRR). Particularly important course policies related to academic integrity are discussed below.

**Collaboration.** Collaboration among enrolled students is allowed and encouraged on homework assignments subject to the condition that every collaborator must make material contributions. Material contributions might include participation in group discussions, critique or presentation of a proposed solution, comparing numerical answers, and the like. However, group submissions are not allowed and you are expected to write up your own work. Copying the work of another student outright, knowingly allowing another student to copy your work, or submitting a copy of a shared set of answers is not acceptable and amounts to a violation of University policy on academic integrity. The best way to adhere to this policy and ensure your collaborations are productive is to:

* attempt problems individually before consulting others
* write up your own solutions in private

Collaboration is not permitted on tests and will result in loss of credit.

**Use of AI.** Learning to use AI effectively and responsibly for problem-solving in an academic context is a skill unto itself. Submitting problem prompts directly to ChatGPT will, most of the time, return superfluous, tangential, and erroneous answers that do not meet assessment criteria for satisfactory work. Furthermore, even when AI-generated material is technically accurate, outputs rarely conform to the examples set forth in class or the solution strategies that you have been taught.

So in the best-case scenario, AI-generated material might be useful but only if you expend additional effort refine the prompts you use and subsequently to parse, understand, and integrate outputs with class content. In the worst-case scenario, AI-generated material will be wrong or irrelevant and simply confuse you. Considering you are learning material that is new to you, you will most likely not be able to distinguish correct from incorrect outputs – if you could, you would have had no need to query in the first place – and it will therefore be difficult if not impossible to use AI effectively. Thus, **using AI is more likely to hinder than to help your learning**, and for this reason I do not recommend it.

Should you choose to use AI you must use it as an aid only and not as a substitute for doing your own work. You will be responsible for using it thoughtfully and judiciously. That means critically assessing any outputs and continuing to prepare work to be submitted in your own words and using your own analyses. Submitting AI-generated outputs directly is never acceptable — doing so amounts to falsely representing material that you did not create as your own work and is a violation of University academic integrity policy. I will respond to such violations as follows.

* some AI-generated content detected: loss of credit and warning
* flagrant AI plagiarism, first offense: loss of credit and report to OSRR
* flagrant AI plagiarism, second offense: automatic course failure and report to OSRR

If you are unsure about where the line is between acceptable and unacceptable use in any particular situation, please discuss the situation with me – I’d much rather help you learn to navigate the issue without the use of penalties wherever possible.

#### Assessments and final grades

I make every effort to provide consistent, fair, and accurate evaluation of student work. Please notify me 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 you feel your grade is unfairly assigned at the end of the course, you have the right to appeal it according to the procedure outlined [here](https://academicprograms.calpoly.edu/content/academic-petition-appeals).

#### Communication and email

Students are encouraged to use face-to-face means of communication (class meetings and office hours) when possible. Every effort is made to respond to email within 48 weekday hours; please be aware that 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.

#### 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](https://drc.calpoly.edu) (DRC).

#### Copyright and distribution of course materials

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.