EPS/ESE 102: Data Analysis and Statistical Inference in the Earth and Environmental Sciences

Fall 2024

Course description:

A practice and application-oriented course covering statistical inference, hypothesis testing, regressions, Monte Carlo methods, time series analysis, and data filtering and visualization. We will also introduce machine learning (ML) methods and Bayesian analysis. The course emphasizes hands-on learning using real data drawn from geophysical and atmospheric science observations. Students will take measurements using phone sensors and provided instruments to reinforce the lecture material and to complete two projects. Coding will be conducted in Python.

Instructors: Roger R. Fu

Email: rogerfu@fas.harvard.edu

Office hours: TBD

Teaching Fellows: TBD

Email TBD

Office hours: TBD

Meetings: Lecture time: Monday and Wednesday. 3:00-4:15 PM.

Lecture location: TBD

Section time: Time TBD

Field trip: One half-day fieldtrip to the Middlesex Fells to take data for the first of two projects. Any student who cannot make this trip must speak with instructor to arrange an appropriate make-up assignment. **Tentatively planned for Saturday, Sep. 14th.**

Prerequisites: Single variable calculus is sufficient, although preparation at the level of Math 21a,b is encouraged. Basic programming experience and some experience with statistics (e.g., mean, standard deviation, linear regression) are useful, but not required.

Grading: 40% Two data acquisition and analysis projects

25% Problem sets

25% Between-class activities (1 per class)

10% Class participation

Late policy: All students will have 5 â€œlate daysâ€ (24-hour delays) to use automatically without penalty. Up to two late days can be used on any one assignment. Late days can't be used on the mid-term or final projects. Beyond these late days, students must request an extension ahead of the due date, or ASAP. Unexcused late work will receive a 10% penalty per day.

Text: None.

Class calendar (subject to minor changes):

Day	Date	Content
Wednesday	9/4	Introduction and overview. Frequentist and Bayesian conception of statistics.
Wednesday	9/11	Same or different? Lecture 1: Explain concept of parameter and talk about mean, standard deviation, and standard error. Assign first problem set. Make groups for first project
Monday	9/16	Roll out First Project. Introduction the Central Limit Theorem
Wednesday	9/18	Same or different? Lecture 2: Introduce t-test
Monday	9/23	Same or different? Lecture 3: Finish t-test for difference.
Wednesday	9/25	Same or different? Lecture 4: Kolmogorov-Smirnov test for difference.
Monday	9/30	Regressions 1: Explain maximum likelihood estimator, weighted mean, and the ordinary least squares regression.
Wednesday	10/2	Regressions 2: Going beyond the OLS: Type II and York regressions.
Monday	10/7	First project cross-pollination day. First problem set due.
Wednesday	10/9	Regressions 3: Finish up York regressions. Overfitting and what to do about it. Multiple regressions. Assign second problem set.
		Resampling techniques 1:

Wednesday	10/16	Bootstrap resampling. Theory and applications.
Monday	10/21	Resampling techniques 2: More bootstrap examples. Start on its big brother, Markov Chain Monte Carlo
Wednesday	10/23	Resampling techniques 3: More examples of MCMC and make your own MCMC code.
Monday	10/28	First project presentations.
Wednesday	10/30	Bayesian inference 1: Bayes theorem, basic examples and intuition, Bayesian conception of parameters and probability.
Monday	11/4	Which model is best? Start on Bayesian model selection. Explain marginal likelihood. Second problem set due. Assign third. Second project rollout
Wednesday	11/6	Bayesian inference 2: Choice of priors, limitations and when is Bayesian inference most appropriate. Second problem set due
Monday	11/11	Practice more with Bayesian model selection. Derive the Bayesian information criterion and understand its uses.
Wednesday	11/13	Time series 1: Smoothing and interpolation using loess, penalized splines, and Savitzky-Golay.
Monday	11/18	Time series 2: Spectral analysis, autocorrelation, and red-shifted noise.
Wednesday	11/20	Time series 3: Fitting red-shifted data and extracting confidence intervals.
Monday	11/25	Machine learning 1: Introduction to concepts and regression problem.
Monday	12/2	Machine learning 2: Components of a convolutional neural network.

Wednesday 12/4	Second project presentations. Third problem set due.