Office Hours

- Friday 5/3: 3-5 pm Lowell
- Sunday 5/5: 8-10 pm Lowell
- Monday 5/6: 3-5 pm Lowell

Final exam

- Tuesday, May 7th, at 9AM - Noon, Science Center Hall C

  ===>>> BRING LAPTOPS! <<<===

- Exam problems will be related to problem sets and in-class problems.
- The exam will focus on material in classes 8-24, Psets 4-10, textbook chapters 7-13 and 15.
- No probability problems (Chapters 2-5), but you will need to use probability concepts in the material listed above.
- Format
  - **PART 1: 2 hours closed book with:**
    - **2 pages of notes 8.5x11 inches, anything written both sides**;
    - Printed **Hypothesis Test Tables** from textbook, which will be attached to the exam. (pdf version is posted in Final Exam module on Canvas)
  - **PART 2: 50 minutes with laptops (MATLAB, regression, ANOVA)**
- During Part 2 of the Final Exam, you can access the internet and files on your laptop (class slides, MATLAB files, etc.).
- During Part 2 of the Final Exam, you will need to use .csv data files to perform analyses, as we did in class sessions. The files will be emailed to you during the exam. Alternatively, you will be able to download files from a Canvas module.

Midterm: Exam with solutions

Course description: Introductory statistical methods for students in the applied sciences and engineering. Random variables and probability distributions; random sampling, including random samples, statistics, and sampling distributions; the Central Limit Theorem; parameter estimation and confidence intervals; hypothesis testing; simple and multiple linear regression. Design of experiments and statistical process control will be covered as time permits.

Prerequisites: Applied Mathematics 21b or Mathematics 21b or equivalent; single-variable calculus will be used frequently, and multivariable calculus and linear algebra will be used occasionally.

Website: https://canvas.harvard.edu/courses/126975 or access from course listing on courses.my.harvard.edu

Teaching Staff

Instructor

Rob Howe howe@seas.harvard.edu

Office Hours: After class and by appointment.

Teaching Fellows

Xiaolong Luo xiaolongluo@fas.harvard.edu

Daniel Purizhansky. dpurizhansky@college.harvard.edu

Caiwei Tian ctian@g.harvard.edu

Administrative assistant

Melissa Majkut mmajkut@seas.harvard.edu

Course Goals

Data in the wild: youâ€™ve got data, from a scientific experiment, from a prototype test, from a marketing survey, from a political poll. What conclusions can you draw, and how strongly? Can you determine cause and effect? How certain can you be that the data average is close to the overall population average? How can you design experiments to draw the strongest inference from the data, and build accurate models of the system?

This is an introductory course for students in engineering and the sciences. It aims to provide tools:

- To become literate in the language of probability and statistics;
- To apply the principles of statistical inference;
- To understand, analyze, and build statistical models.
- To use statistical methods to design effective and efficient experiments

Similar courses: This course covers the material in Statistics 110 (Probability), 111 (Statistical inference), and 139 (Statistical modeling), but at less depth, with less theory, and with an emphasis on applications. Engineering Sciences 150 and Statistics 110 cover probability but not statistics or modeling.

Class Format

This will be a â€œflipped classroomâ€ course. Readings in the textbook and/or watching the lecture videos are required in advance. Short summary lectures will be presented at the top of the class. Class time will focus on in-class investigations, computer simulations, and small-group problem solving and data analysis.

Please bring your computer with you to each class!

Sections

There will be one section offered each week. These will aid you in understanding course material and solving homework problems. These are not mandatory but are highly recommended. The times/locations will be determined after a poll which will be sent out during the first week, probably Monday afternoon or evening.

2.13 Section Note: PDF
2.20 Section Note: PDF
4.02 Section Note: PDF
4.09 Section Note: PDF

Homework Policy

The assignments will give you practice working with the course concepts, and hence to prepare you for the exams â€“ and eventually, for use in your professional lives. We encourage collaboration and discussion of the course material; however, violations of academic integrity (such as direct copying of

othersâ€™ work) will be sanctioned. ***You must write up your own problem sets and acknowledge those you worked with***.

Late assignments will only be accepted within a day of the deadline and will receive a 25% penalty to the grade, unless excused in advance by the instructors.

PSet1: PDF | Solution
PSet2: PDF | Solution
PSet3: PDF | Solution
PSet4: PDF | Solution
Pset5: PDF | Solution
Pset6: PDF | Solution
Pset7: PDF | Solution
Pset8: PDF | Solution
Pset9: PDF | Solution
Pset10: PDF | Solution

Evaluation

- Weekly problem sets (about 10 total) â€” 25%
- Project on regression modeling (during reading period) â€” 15%
- One midterm exam (March 7, in class) â€” 25%
- Final exam (May 7, 9:00 am â€“ Noon, location TBA) â€” 35%

Required Text

Applied Statistics and Probability for Engineers, Douglas C. Montgomery, George C. Runger (Wiley). You can use the 4^{th}, 5^{th}, 6^{th}, or 7^{th} edition, which are only slightly different from each other. Note that the earlier editions may be available at a lower price.

You can access the full text of the 5^{th} edition on-line through the Harvard Libraries:

https://hollis.harvard.edu/permalink/f/b2fb8a/01HVD_ALMA512229467170003941

MATLAB Software

We will be using MATALAB for statistical calculations in class as well as for the problem sets and the project. See instructions on downloading MATLAB to your computer: **http://downloads.fas.harvard.edu**

(use the MATLAB key provided on this webpage for your platform â€“ Mac or Windows). You can use the online MATLAB tutorials: https://www.mathworks.com/help/matlab/getting-started-with-matlab.html

MATLAB Boot Camp: If this is your first time with computing and/or MATLAB we strongly recommend that you attend the MATLAB Boot Camp that offers self-paced lessons with office hours during the first week of the semester. More information and registration is available at https://canvas.harvard.edu/courses/78252

Course Topics

- Data Summary and Presentation. Definition of a random sample and descriptive statistics.
- Random Variables and Probability Distributions. Random experiments; axioms of probability. Conditional probability and independence. Random variables. Discrete and continuous distributions. Expected value and variance. Common distributions, including the normal, binomial, and uniform. The Central Limit Theorem and sampling distributions.
- Decision Making for a Single Sample. Point estimation of parameters. Estimator bias, variance, and Formulation of statistical hypotheses. Definition of a hypothesis test and associated terminology. Hypothesis tests for typical cases (normal distribution, t, chi-squared, etc.). Type II error and the choice of sample size. Definition of a confidence interval. Confidence intervals for typical cases.
- Decision Making for Two Inference on the means of two populations (variance known or unknown) and two population proportions. Paired t-tests.
- Building Empirical Models: Simple and multiple linear Hypothesis tests and confidence intervals in regression models. Adequacy of the regression model: model checking and residual analysis.

- Design of Engineering Experiments (time permitting). Factorial experiments. Blocking. Model adequacy checking.
- Statistical process control (time permitting). Control chart design and use.