

Introduction to Probability and Statistics for Computer Science Majors

37755 Stat 67 Lecture B

1pm-1:50pm MWF HH178

Discussion1 F 11-11:50 am SSL 248

Discussion2 F 12-12:50 am SSL 248

Course Website: <https://canvas.eee.uci.edu/courses/1744> Visit the course website to find a google calendar, course objectives, assignments, slides, worksheets, office locations, office hour times, and more!

Instructional Team

Wendy Rummerfield-
Discussions/TA

Help with homework and
assignments, content questions

wrummer@uci.edu

Yannan Tang-Grader

Questions about grading, how
to do problems you missed, pick
up assignments

yannant@uci.edu

Elizabeth Ben Ward -Lecturer

Questions about content, help
with assignments, questions
about test grading,
administrative matters

eward@uci.edu

Come to **office hours** (times and locations on website): for help with your homework or project, so that I know who you, so that the TA knows your name, for questions about content, just to say “hi”, to steal candy from the candy bowl, to ask the grader why you got a certain grade, and to pick up work.

Textbooks: Diez , Barr, Cetinkaya-Rundel “OpenIntro Statistics”, 3rd edition, Creative Commons license. Visit openintro.org for a free pdf or amazon.com for printed edition.

Olofsson & Andersson “Probability, Statistics and Stochastic Processes” 2nd Edition, Wiley, Sections 2.1, 2.3, 2.5.4, 2.6, 2.7, 3.9, 6.4.2, and 6.8. Other sections may be useful as well. This book is available through the library in pdf form. You can use a VPN or a campus computer to search for the book on the library website, then click on “ebook”.

Discussion Section: Students must go to their assigned discussion section unless you have permission from Wendy. Wendy will be giving lectures, practice problems and doing activities in discussion section.

Preferred name policy: If you prefer a name other than your given name, or would prefer not to share your name with the class, choose a different name we may address you by. Please write your name on the top sheet of all your papers in the top right corner. **Put your student ID number at the bottom right corner of your paper not at the top.**

Grading:

Clicker Questions/Weekly Quizzes: 20%

Test 1: 20%

Test 2: 20%

Project 1: 10%

Project 2: 5%

Final Project: 15%

Homework: 10%

Tentative dates for **Tests** are Wednesday in weeks 6 and 10. If you will be unable to attend class on the day of a test, please email in advance (at least 24 hours) so that we can make arrangements. If you have an emergency, please email me as soon as you know you will not be at the test.

The **final** will be a project assigned 2 weeks before the end of term and due during our courses' final exam time. That is Wednesday, June, 8th at 1:30pm. You may turn it in early.

Clicker Questions will begin in week 2. Each question will be worth two points, one for participation and one for getting the correct answer. At the end of each week an online **Quiz** will be posted with questions similar to the clicker questions. You will get to keep the score for either the quiz or the clicker questions, whichever you score higher on. If you miss class or forget your clicker, you will have the quiz to make up for those lost points.

There are **homework** assignments, each should take 2 hours or less. Homework will be available in advance on the class website. Homework is due before the start of class. You may turn it in during class or to the box on the 2nd floor of DBH any time before class begins. The lowest 2 homework scores will be dropped.

Homework must be neat, clearly written and easily readable. Your preferred name should be in the upper right corner of the first page. Your student ID should be in the lower right corner of the paper. Please write/print only on one side of the paper. The problems should be in order. If your homework is two or more pages long, the pages should be stapled together. If the grader determines that your homework is not easily readable, you will be given a zero. Yannan will be grading hundreds of papers per week, please respect his time and effort by ensuring that your homework is done neatly. Not every problem will be graded for correctness, some may be graded for completion.

Homework scores:

- 5-All problems complete, neatly presented, graded problems correct.
- 4-All problems complete, neatly presented, graded problems mostly correct.
- 3- Something small is missing, graded problems mostly correct.
- 2- Graded parts incorrect or missing whole problems.
- 1-Missing two or more problems or much of the graded work is incorrect.
- 0-Not turned in

Computer based assignments: There will be quite a few times you will be expected to use a computer to run simulations, create graphics and analyze data. You will need a laptop occasionally during class and you will be told in advance which days a laptop is required. I will be using R and R Studio in class. R is a free program and a statistical programming language. R Studio is a free wrapper for R. It will be nice to know R if you plan on doing any data analysis, data mining or predictive modeling in the future. It is a great language for prototyping. You may use any computer software you would like, but I won't be able to help you and the assignments will be based in R.

We will be using The Island, a simulated population of people, to complete some projects and assignments. A unique login will be provided for each student. The best way to learn statistics is by doing statistics. The Island allows us to design and run studies in a timely fashion while still facing some of the difficulties of sampling and experimental design.

LaTeX is a typesetting language that allows for an easy way to type math and gives documents a professional look. Now is a great time to learn LaTeX. It will make your projects or homework look really nice. If you plan on grad school or want to impress your professors, this is a great tool to have.

The scale conversion to letter grades will be:

- A+ ≥ 98 > A ≥ 92 > A- ≥ 90
- B+ ≥ 88 > B ≥ 82 > B- ≥ 80
- C+ ≥ 78 > C ≥ 72 > C- ≥ 70
- D+ ≥ 68 > D ≥ 62 > D- ≥ 60 > F

The instructor reserves the right to adjust the grading scale in the direction which favors the students.

Students who get an A will show mastery of the objectives on tests and projects. They will turn in complete assignments, done neatly and correctly. During class, they will make the effort to participate. They will take the opportunity to learn the things they missed and show their understanding.

Class behavior: Refrain from packing up or leaving early. Do not hold side conversations while the instructor or your classmate is speaking. Avoid texting, reading the news, playing hearthstone, and checking email or facebook during class.

Academic dishonesty: Copying homework from a classmate or a solution key is considered cheating. Students may work together on homework, but must write their answers independently. Looking at another student's test, looking at a cheat sheet or phone, talking or texting about the test while it is in progress are all considered cheating. Projects should contain only your work. Any text from websites, academic papers or books must be properly cited. Others may edit your draft, but under no circumstances should anyone be doing the analysis or the writing other than yourself. Any code you use must be written by you. If the rules are broken, consequences will follow as laid out in the UCI Academic Senate Policy on Academic Honesty.

Topic 1: Probability (Week 1)

Definition of probability, events, union, intersection, venn diagrams, mutually exclusive (disjoint), conditional probability, independence, Bayes rule, hypothetical hundred table

Topic 2: Study Design (Week 2)

Survey, experiment, observational study, common mishaps in sampling, randomization, block design, representative samples, stratified sampling, association vs. causation

Topic 3: Monte Carlo Methods/Simulation (Week 2-Week 10)

Simulation for combinatorics (week 2), probability (week 2), random variables (week 4), Bayesian estimators (week 5), sampling distributions (week 6)

Topic 4: Summarizing Data (Week 3)

statistical graphics (using computer), tables, summary statistics, numeric variables, categorical variables, sample mean, sd, median, IQR, outliers, correlation, contingency tables, histograms, boxplots, bar charts, scatterplots, line of best fit (linear regression)

Topic 5: Random Variables (Week 4)

Discrete random variables, continuous random variables, probability using distributions, expectation, mean of X , variance of X , quantiles of X , normal, binomial, Bernoulli, exponential, gamma, beta, bivariate normal, other distributions as needed.

Topic 6: Bayesian Estimators (Week 5) With Guest Lecturer Fletcher Christensen!

Choosing a prior, posterior (for conjugate situations or using Bayes rule), credible intervals, posterior functionals as estimators

Topic 7: Frequentist Estimators and Hypothesis Tests (Week 6 –Week 9)

sampling distribution for small n , sampling distribution for a mean, sampling distribution for a proportion, CLT, t distribution, confidence intervals, hypothesis tests (one mean, two means, proportion, two proportions, test of slope in regression, chi squared, ANOVA), type 1 and type 2 error, multiple comparisons, spurious correlations, practical vs. statistical significance, p values, null and alternative hypothesis

The above schedule, policies and assignments in this course are subject to change in the event of extenuating circumstances or by mutual agreement between the instructor and students.