ECON 57 - Introduction to Statistics and Probability

Fall 2023
Augusto Gonzalez Bonorino

September 8, 2023

Course Information (Section 02)

Office Hours: Monday any time before 11am or after 4:30pm by appointment; Friday's 10am to 1pm (office Carnegie 201, zoom by appointment)

E-mail: agxa2023@pomona.edu

Mentor: Eric Li (eylq2022@mymail.pomona.edu)

Lectures: Tuesday, 09.35 am – 10.50 am; Thursday, 09.35 am – 10.50 am (Hahn Social Science Bldg 101, Pomona College)

Room: Hanh 101

Prerequisites: MATH 30 or equivalent

Department of Economics – Statement about Email and Appointments. Because we function as a community where we all have many demands on our time, respecting the time of others is crucial. To this end, all members of the Department of Economics community need reasonable expectations about how time is allocated outside of class. Faculty try to give students ample advance notice about deadlines and to honor times when students have extracurricular activities. We ask that students extend the same courtesy to faculty, who also have additional demands on their time. So, if you email us, expecting an instantaneous response is not reasonable. Instead, expect a response within a few days. However, please note that response times can be longer over weekends and holidays. If it's an emergency, faculty will try to get back to you sooner. A polite reminder follow-up is a good idea if the professor doesn't respond in a few days. Office hours are the usual time when faculty are available to answer student questions. However, if you absolutely cannot make office hours, you may ask for an individual appointment and we will try our best to accommodate you. Note, though, that this may not always be possible. Thank you for your understanding.

Course Description

ECON 57 is a course in basic probability and statistics that consists of three major segments: descriptive statistics, probability theory, estimation and inference. Examples will be often used to support in-class discussions and theoretical explanations. In addition, this course complements the theoretical component with an introduction to modern data science with R programming. This fourth component is intended to prepare students for the modern workforce and provide them with computational skills that will prove valuable for any data-oriented job. The ultimate goal of this course is for the students to master the essential toolbox that they will need to

learn and apprehend the material covered in upper-division economics courses and to become effective self-directed learners.

Course Requirements

There will be a total of 2 presentations per group, 2 Exams, 4 homeworks, 4 R programming labs, and a group research project.

There will be two take-home exams throughout the course. There is no final exam, you will have a research project instead. To be excused, a student must pre-notify the instructor before an exam by e-mail. Missing an exam without pre-notification and written verification will result in a flunking grade for that exam. Exceptions will be considered and exemptions granted only in serious medical cases, family emergencies, or other extreme circumstances, which must be documented properly. **there will be no make-up exams**. A student who has been properly excused from an examination will receive a final grade based on the other exams in the course.

- Group Presentations (2 per group): 15%
- Labs (4): 15%
- Homeworks (4): 20%
- Exams (2): 20%
- Group Project: 30%

Assignments Instructions

Presentations: These are your opportunity to share key insights you and your team members have gathered from previous lectures with your classmates, an unorthodox idea about topics covered until then, expand on covered topics, or showcase something cool you built (again, must be related to one of the topics covered until then). On the first week of class, we will create the groups and randomly assign a week to present. This way everyone knows when they must present with enough time in advance to prepare a quality presentation. I will assess the students individually and as a team. Your final grade for the presentation will be an average of your individual and team grade. The following will be assessed:

- Quality of presentation. Were your slides readable? Was the information communicated clearly?
- Time. Presentations must be around 8 to 12 minutes long.
- Insightful. Did we learn something new from your presentation? Are your ideas original?

Some sample projects from previous years include mapping historic events related to statistics to current economic or political events, a UI interface to study simulations, animations of statistical operations, visualizations of notable distributions, and applications of covered topics to current economic topics.

Homeworks: All homeworks will contain 7 to 10 exercises that cover topics from the previous two weeks. These will be a mix of conceptual and mathematical exercises to help you 1) practice solving problems, 2) develop an intuition of statistics and probability, and 3) apply the ideas to economic scenarios. You can collaborate with your classmates or team members, but every student is expected to submit their individual work. If you do collaborate, please write

down the names of those you work with at the beginning of the document. Each homework is worth 5%.

Labs: The labs are intended to test and push your R programming skills. They are a small part of your grade but will challenge you to 1) apply methods covered in class and 2) research new tools to solve unseen problems. Each lab will consist of 7 to 10 coding exercises of varying difficulty. You must create and submit an R markdown notebook for each lab.

Exams: The first one will test your knowledge of descriptive statistics and basic R programming, while the second exam will test you on probability theory and probabilistic programming in R. Both will be take-home exams. Exams will start on that week's Thursday at 1pm and be due the next day at 11:59pm. This gives you 36hs to complete your exam. The final submission should be a PDF document or an Rmarkdown notebook with your solutions to both mathematical and programming exercises.

Final Project: The final project is an opportunity to develop original work. You can work as a team or individually. Regardless if you are interested in a research or corporate job or a life of entrepreneurship, the ability to analyze data and produce original work will set you aside from the average applicant or competitor. This final project is your chance to start working, formally, on an idea you have held for a long time. This project must showcase your ability to leverage R programming for data analysis and simulations, and your capacity to apply the concepts learned in class to conduct economic analysis.

Course Material

All of the necessary material for the class will be available to you through Canvas. If you would like to dive deeper, I will host the course's and additional material in the following GitHub repository. Here, you will find the lecture notes, slides, R notebooks, papers, textbooks, and additional resources.

While there is no required textbook for the class, here are some recommendations you can find at the library plus online resources:

- Statistics for Business and Economics
- Naked Statistics
- Probability
- Learning base R
- R for statistics
- The Book of Why

Weekly Schedule

Week 1 (T: 08/29 - Th: 08/31)

- Syllabus
  - Course content & structure, grading, homeworks, and exams.
- Brief history of statistics, Population vs Sample, Random Sampling.
- Parameter vs Statistic.

- Statistical data types.
- R data types.
- Introduction to R programming - Part I,
  - Installing R and Rstudio,
  - R notebooks and Markdown.
  - Variables
  - R data types (integer, float, string, boolean, vector, factor).
  - Conditionals
- **Start:** Essay on history of statistics.
- **Due:** Fill out course questionnaire.

Week 2 (T: 09/05 - Th: 09/07)

- Data, Information, & Statistics.
- Frequency Distributions/tables (only for univariate case)
  - Discrete type
  - Continuous type
- Measures of centrality
  - Mean,
  - Median,
  - Quantiles,
  - Mode.
- Measures of dispersion
  - Variability or dispersion,
  - Indices of dispersion (variance, standard deviation, and coefficient of variation).
- Skewness and Kurtosis.
- Introduction to R programming - Part II,
  - Matrices and tables.
  - Frequency distributions
  - Descriptive statistics
- **Start:** HW #1 on weeks 1 & 2.
- **Due:** Essay on history of statistics.

Week 3 (T: 09/12 - Th: 09/14)

- Presentation #1
- Statistical thinking,
  - How to think about distributions & statistics.
- Frequency and Density function
  - Discrete
  - Continuous
- Cumulative Frequency function
- Visualizing frequency tables and data.
- Functions and loops in R
- Descriptive statistics in R
- Comprehensive exercises
- **Start:** Lab #1 on weeks 2 & 3.
- **Due:** HW #1 on week 1 & 2.

Week 4 (T: 09/19 - Th: 09/21)

- Presentation #2.
- Bidimensional distributions
- Joint frequency tables.
- Absolute vs Relative vs Conditional frequencies
- Statistical independence.
- Covariance & Correlation
- Reading and Writing files in R.
- Two-way entry table in R
- Covariance and correlation in R.
- **Start:** HW #2 on weeks 3 & 4.
- **Due:** Lab #1 on weeks 2 & 3.

Week 5 (T: 09/26 - Th: 09/28)

- Presentation #3.
- Recap of descriptive statistics and distributions.
- Frequentist vs Bayesian thinking
- A conceptual introduction to statistical and causal inference.
- Advanced visualizations in R with ggplot2
- Data wrangling with dplyr - Part I.
- **Take-Home Exam 1** on weeks 1 to 4
- **Start:** Lab #2 on weeks 4 & 5.
- **Due:** HW #2 on weeks 3 & 4.

Week 6 (T: 10/03 - Th: 10/05)

- Presentation #4
- Conceptual introduction to probability.
- Randomness
  - How to think about random stuff.
- Set theory
  - Notable sets.
  - Common operations & properties
- Random Experiments, Events, & Sampling
- Counting
  - Multiplication rule
  - Permutations
  - Combinations
- Set theory in R
- Counting in R
- **Start:** HW #3 on weeks 5 & 6
- **Due:** Lab #2 on weeks 4 & 5.

Week 7 (T: 10/10 - Th: 10/12)

- Presentation #5
- Definitions of Probability.
- Kolmogorov axioms
- Conditional probability
- Total probability theorem.
- Calculus review
- Bayesian statistics/probability
- Bayes theorem and Bayes rules.
- Introduction to probability in R - Part I
  - Basic probability
  - Kolmogorov Axioms
  - Conditional probability
- Basic calculus in R (extra)
- Basic Bayesian statistics in R (extra)
- **Start:** Lab #3 on weeks 6 & 7.
- **Due:** HW #3 on weeks 5 & 6.

Week 8 (T: 10/17 - Th: 10/19)

- Presentation #6
- Probabilistic thinking
- Random variables.
- Probability Density Function (PDF)
  - Discrete.
  - Continuous.
- Cumulative Density Function (CDF)
  - Discrete.
  - Continuous.
- Expected value.
- Moments.
- Introduction to probability in R - Part II.
- **Start:** HW #4 on weeks 7 & 8
- **Due:** Lab #3 on weeks 6 & 7.

Week 9 (T: 10/24 - Th: 10/26)

- Presentation #7
- Poker, Machine Learning, and Economics applications
  - Probability distributions
  - Probabilistic programming - modeling distributions
- Mathematical introduction to probability distributions
- Probabilistic programming in R
- **Start:** Lab #4
- **Due:** HW #4 on weeks 7 & 8

Week 10 (T: 10/31 - Th: 11/02)

- Presentation #8
- Gaussian Distributions
- Central Limit Theorem (CLT) & Law of Large Numbers (LLN)
- Exercises with probability distributions
- R programming review - applying what we have learnt
  - Exploratory Data Analysis (EDA)
  - Poker and Economics simulations
- **Take-Home Exam 2** on weeks 5-9
- **Start:** Project Proposal draft
- **Due:** Lab #4 on weeks 8 & 9

Week 11 (T: 11/07 - Th: 11/09)

- Presentation #9
- Statistical inference & Expectation
- Point estimation
- Interval estimation
- Confidence intervals
- Introduction to statistical inference in R
- **Start:** Final project proposal; EDA
- **Due:** Project proposal draft

Week 12 (T: 11/14 - Th: 11/16)

- Presentation #10
- Recap of distributions and estimation
- Hypothesis Testing
- Type I & Type II errors
- Power function
- Test statistics and p-value
- Size of a test
- Hypothesis testing in R
- Review of probability in R
- **Start:** Final R programming review; Introduction, Literature Review, & Data sections.
- **Due:** Final project proposal

Week 13 (T: 11/21 - Th: 11/23)

- Linear Regression
- Introduction to Econometrics
- Advanced data wrangling in R
- Introduction to regression analysis in R
- **Start:** No more
- **Due:** Literature Review, & Data sections of paper; EDA

Week 14 (T: 11/28 - Th: 11/30)

- Concluding thoughts
- Econometrician as Data Scientist
- Custom review for final exam
- Special topic (optional lecture)
  - Special topics: Web scraping in R, Text as (econ) data, Basics of Natural Language Processing (NLP)
- **Start:** Studying for final exam and exploring data with R
- **Due:** Full paper (either draft or final version)

Week 15 (T: 12/05 -

- Final exam
- **Due (end of week):** Full project (paper + code)