STAT 5600-6600:
PROBABILITY AND STATISTICS FOR DATA SCIENCE
Course Syllabus

CONTACT INFORMATION:
Instructor   Elvan Ceyhan
Office   203 Extension Hall
Office Hours   Tuesday 12:00 pm -01:00 pm or by appointment
Phone   (334) 844 3613
E-mail   ceyhan@auburn.edu

TIME & PLACE: Tu-Th 3:30 pm - 4:45 pm ACLC (Academic Classroom/Lab Complex) 010

Recommended Texts:

Probability and Statistics for Data Science: Math + R + Data (Chapman & Hall/CRC Data Science Series) 1st Edition, by Norman Matloff.

All of Statistics: A Concise Course in Statistical Inference by Larry Wasserman (Springer).

Additional Recommended Texts:

Probability and Statistics for Data Science by Carlos Fernandez-Granda. This book is freely available online at Carlos Fernandez-Granda's website.

Statistical Inference, 2nd ed. (Cengage Learning) by Casella, G. and Berger, R. L. (2002).

Mathematical Statistics with Applications in R, 3rd ed. (Elsevier) by Ramachandran, K.M. and Tsokos, C.P.

PREREQUISITE: STAT 3610 or STAT 3010 or MATH 1620 (Minimum Grade of C in any one of them)

COURSE WEBSITE: posted under *Canvas*

SOFTWARE: The *R* software will be used for some simulation studies. *R* can be downloaded for free from http://www.r-project.org or *RStudio* from https://www.rstudio.com

HOMEWORK: At least one homework will be assigned at the conclusion of each chapter.

EXAMS (tentative): There will be two midterms and one final exam. MT1 is scheduled to be in late September-early October and MT2 scheduled to be on early November; both MTs will be during the class time in the regular classroom. One 8.5×11 formula sheet (both sides) is allowed for each midterm (MT) exam and two such sheets (both sides for each) are allowed for the final exam.

GRADING: Homework (20%), 2 MTs (25% each), Final (30%), Project (10% bonus, if assigned).

COURSE CONTENT (provided time allows):

Probability review: Basics: sample space, outcomes, probability; Events: mutually exclusive, independent; Calculating probability: sets, counting, tree diagram; Conditional probability; Law of total probability; Bayes' theorem.

Random variables: Mean, Moments, Variance; pmf, pdf, cdf; Discrete RVs: Bernoulli, Binomial, Geometric, Indicator; Continuous RVs: Uniform(a, b), Exponential(λ), The Normal distribution; Joint distributions & conditioning, Joint probability distribution; Linearity (and product) of expectation, Conditional expectation, Sum of a random number of RVs

Probability inequalities: Markov's Inequality, Chebyshev's inequality, Weak Law of Large Numbers, Central Limit Theorem

Markov chains: Stochastic processes, Setting up Markov chains, Balance equations.

Non-parametric inference: Basics of inference, Simple examples, Empirical PMF, Sample mean bias, se, MSE; Empirical Distribution Function (or eCDF), Kernel Density Estimation (KDE), Statistical Functionals, Plug-in estimator.

Confidence intervals: Percentiles, quantiles, Normal-based confidence intervals, DKW inequality

Parametric inference: Consistency, Asymptotic Normality, Basics of parametric inference, Method of Moments Estimator (MME), Properties of MME, Basics of MLE, Maximum Likelihood Estimator (MLE), Properties of MLE,

Hypothesis testing: The Wald test, t-test, Kolmogorov-Smirnov test (KS test), p-values, Permutation test

Bayesian inference: Bayesian reasoning, Bayesian inference, Priors, Conjugate priors

Regression: Simple Linear Regression, Multiple Linear Regression
Time Series Analysis: EWMA Time Series modeling, AR Time Series modeling

Credit Hours: This course is a 3-credit course, which means that students are expected to do at least 9 - 9.5 hours of course-related work or activity each week during the semester. This includes (where applicable) scheduled class lecture/discussion meeting times as well as time spent completing assigned readings, studying for tests and examinations, participating in lab sessions, preparing written assignments, and other course-related tasks.

Classroom Policies

- **Policies on Class Attendance, Submission of Late Written Assignments, Missed In-Class Work and Missed Examinations:**

  o **Excused Absences:** Students are granted excused absences from class for the following reasons: Illness of the student or serious illness of a member of the student's immediate family, death of a member of the student's immediate family, trips for student organizations sponsored by an academic unit, trips for University classes, trips for participation in intercollegiate athletic events, subpoena for a court appearance and religious holidays. Students who wish to have an excused absence from this class for any other reason must contact the instructor in advance of the absence to request permission. The instructor will weigh the merits of the request and render a decision. When feasible, the student must notify the instructor prior to the occurrence of any excused absences, but in no case shall such notification occur more than one week after the absence. Appropriate documentation for all excused absences is required.

  o **Make-Up Policy:** We jointly will decide to take the proper action and/or make up missed major examinations (e.g., hour exams, mid-term exams) due to properly authorized excused absences. Except in unusual circumstances, such as continued absence of the student or the advent of University holidays, a make-up exam will take place within two weeks from the time the student initiates arrangements for it. Except in extraordinary circumstances, no make-up exams will be arranged during the last three days before the final exam period begins. The format of the make-up exam will be (as specific by the instructor).

- **Email and Canvas Use:** Students have control of the notification settings in their Canvas accounts. You should set up their notifications to alert them when an Announcement is posted, an Assignment is due, a grade is released, etc. For students new to Canvas, save time (and emails) by sharing this link to a 7 minute ["Getting Started with Canvas" video (and transcript) (Links to an external site.)](#) created by Canvas LMS.

- **Provisions of the Americans with Disabilities Act:** Students who need accommodations are asked to electronically submit their approved accommodations through AU Access and to make an individual appointment with the instructor during the first week of classes – or as soon as possible if accommodations are needed immediately. If you have not established accommodation through the Office of Accessibility, but need accommodation, make an appointment with the Office of Accessibility, 1228 Haley Center, 844-2096 (V/TT).

- **Academic Honesty:** All portions of the Auburn University Student Academic Honesty code (Title XII) found in the [Student Policy eHandbook](#) will apply to this class. All academic honesty violations or alleged violations of the SGA Code of Laws will be reported to the Office of the Provost, which will then refer the case to the Academic Honesty Committee.

- **Classroom Behavior:** The Auburn University Classroom Behavior Policy is strictly followed in the course; please refer to the [Student Policy eHandbook](#) for details of this policy and the [Policy on Classroom Behavior](#).

- **Emergency Contingency Statement:** If normal class and/or lab activities are disrupted due to illness, emergency, or crisis (such as an H1N1 flu outbreak), the syllabus and other course plans and assignments may be modified to allow completion of the course. If this occurs, an addendum to your syllabus and/or course assignments will replace the original materials.

Additional Policies

Policies Related to Generative Artificial Intelligence Tools

AI Policy: Permitted in this Course with Attribution: In this course, students are allowed to use Generative AI Tools like ChatGPT to support their work. To maintain academic integrity, students must disclose any AI-generated material they use and properly attribute it, including in-text citations, quotations, and references (such attribution and acknowledgement is also required any other online tool or book or written work). Students should exercise caution and avoid sharing any sensitive or private information when using these tools. Examples of such information include personally identifiable information (PII), protected health information (PHI), financial data, intellectual property (IP), and any other data that might be legally protected.

A student should include the following statement in assignments to indicate use of a Generative AI Tool: "The author(s) would like to acknowledge the use of [Generative AI Tool Name], a language model developed by [Generative AI Tool Provider], in the preparation of this assignment. The [Generative AI Tool Name] was used in the following way(s) in this assignment [e.g., brainstorming, grammatical correction, citation, which portion of the assignment]."

Policies Related to COVID-19

- Due to the Coronavirus pandemic, public health measures have been implemented across Auburn's campus. Students should stay current with these practices and expectations throughout the campus. For the latest information about Auburn University's plans for the

Fall 2021 semester, please visit AU's [COVID-19 Resource Center website](#), which will include updates as they are made available. The resource center website offers everything from campus messages to information on policies as well as helpful resources and details about the COVID-19.

- **Face Coverings:** We will follow the Auburn University requirements on face covering indoors, although please be mindful of the others, and use face covering if you are feeling or having the symptoms or have been near someone who is sick or showing the symptoms. If face covering policy is in effect, students in this class are required to wear face coverings that appropriately cover the nose and mouth to limit the spread of infectious disease. Failure to comply with the requirement represents a potential Code of Student Conduct violation and may be reported as a non-academic violation. Please consult the [Policy on Classroom Behavior](#) for additional details.