



THE OHIO STATE UNIVERSITY

STAT 3201

Introduction to Probability for Data Analytics (24996)

Autumn 2026

Syllabus - Instructor Information

Kai Tan

Professional Qualifications

Ph.D., Statistics, Rutgers University

Master's, Statistics, University of Kentucky

Master's, Statistics, East China Normal University

Bachelor's, Mathematics, Central China Normal University

Syllabus - Required or Recommended Readings

Required textbook: *Mathematical Statistics with Applications* (7th edition), by Wackerly, Mendenhall, and Sheaffer

The textbook will be supplemented with additional material. For computing with R, the following free resource will prove useful:

Introduction to Probability and Statistics using R, by Kerns; available online at <https://ipsur.r-forge.r-project.org/book/download/IPSUR.pdf>

Syllabus - Course Schedule

Below is a **tentative** weekly schedule, subject to change as the class progresses. Note that chapter references refer to the textbook "Mathematical Statistics with Applications".

Week	Topic
1	Course introduction. Introduction to R (Chapters 1.1 - 1.2). 2 lectures; class begins on Wednesday.
2	More on R. Summarization of data – graphical and numerical summaries. From summarization of a sample to inference about a population (1.3-1.6). Basics of probability. Definitions, formal mathematics for probability (2.1-2.2). 3 lectures.
3	Probability models for discrete data. Calculating a probability (2.3-2.5). 2 lectures. Monday is Labor Day, no class.
4	Counting methods. Conditional probability (2.6-2.7). 3 lectures.
5	Bayes' rule. Strategies for computing probabilities. Examples (2.8-2.13). 3 lectures.
6	Discrete random variables. Probability mass function, mean and variance. Bernoulli distribution (3.1-3.3). 3 lectures.
7	Midterm 1. Named discrete random variables. Binomial distribution. Geometric distribution. Negative binomial distribution. Hypergeometric distribution. (3.4-3.7). 3 lectures; one lecture devoted to the exam. [Midterm 1, Monday, Oct 5]
8	Poisson distribution. Continuous random variables. Probability density function. Computing probabilities (3.8; 4.1-4.2). 2 lectures; Friday is Fall Break, no class.
9	Continuous random variables – mean and variance. Named continuous distributions. Uniform distribution. Normal (Gaussian) distribution (4.3-4.5). 3 lectures.
10	Named continuous distributions. Exponential distribution. Gamma distribution. Beta distribution (4.6-4.7). 3 lectures.

Week	Topic
11	Functions of random variables. Method of distribution functions. Method of transformation (6.1-6.4). 3 lectures.
12	Catch up. Midterm 2. 2 lectures; one lecture devoted to the exam. Wednesday is Veteran's Day. [Midterm 2, Friday, Nov 13]
13	Sampling distributions. Simulations. The central limit theorem (7.1-7.3). 3 lectures.
14	Bivariate probability distributions. Computing probabilities. Marginal and conditional distributions (5.1). 1 lecture; Wednesday and Friday are Thanksgiving Break, no class.
15	A random sample (independent random variables) (5.2-5.4). 3 lectures.
16	Covariance and correlation. Conditional expected values (5.7, 5.11). 2 lectures.
Final Exam	As scheduled by the registrar. In the regular classroom. Thursday, December 17, from 10:00 am to 11:45 am.

Course Details - Instructor Contact Details

Instructor email address: tan.1457@osu.edu

- Use your OSU name.# email for all correspondence, email from a non-OSU account will not be answered.
- Please begin subject with "STAT 3201" and avoid using the Carmen email tool.
- I will attempt to answer emails within 2 business days (excluding weekends and university holidays). If you haven't heard back within that window, please re-send your message in case it was missed.
- Many questions, especially those about homework, are better answered in person. Feel free to stop by my office hours or to make use of the tutoring room ([detail will come soon](#)).
- For questions that come in via email, if the question is of general interest, I may post your question verbatim on Carmen Discussions section, along with a response for all to see.

Course Details - Office Hours

Office hours: Wednesday 10 - 11 AM, 2 - 3 PM, and by appointment

Office location: Cockins Hall, 205C

Course Details - Additional Instructional Staff

The grader for this course is **Dayeon Jung**, a graduate student in the Statistics Department. Her email address is jung.1017@osu.edu. For all homework grading questions, contact the grader.

Course Details - Class Meeting Details

Class Meeting Details: 11:30 AM-12:25 PM - Jennings Hall 060 - MWF

Course Details - Course Description

An introduction to probability and its role in statistical methods for data analytics. Equal emphasis is placed on analytical and simulation-based methods for quantifying uncertainty. Approaches to assessing the accuracy of simulation methods are discussed. Applications of probability and sampling to big-data settings are discussed. Prereq: Math 1152, 1161.xx, 1172, 1181, or equiv; or permission of instructor. Not open to students with credit for 4201 or Math 4530.

Prerequisite(s) / Corequisites: Prereq: Math 1152, 1161.xx, 1172, 1181, or equiv; or permission of instructor. Not open to students with credit for 4201 or Math 4530.

Course Details - More About This Course

This course serves several audiences -- the majors in Statistics and in Data Analytics, the minor in Statistics, and those who simply want to know a bit about probability and statistics. It is the first substantive course in the minor/major and

is part of a year long sequence. This semester focuses on probability, from the basic definition and historical problems, to discrete and continuous random variables, to varied techniques useful for summarizing and working with random variables, and beyond. Along the way, we will learn about parallel summarizations of data sets.

Paper and pencil work: Much of the material in this course can only be internalized by reading the text, participating actively in lectures, and working problems "by hand". Lectures and homework assignments are designed to help you build the skills that will be helpful in later courses and in your post-collegiate life! Investing time and effort now will be worthwhile in the long run.

Computing: The course will make use of the R computing environment. R is a statistical software package that is available for free download ([link](#)). I strongly encourage you to make use of RStudio ([link](#)), as this platform will be used heavily in future courses. This course will get you the tiniest start on R. You will use R more in subsequent courses in the minor and majors.

Course Details - Course Learning Outcomes

By the end of this course, students should successfully be able to:

1. Quantify uncertainty about events using mathematical descriptions of probability
2. Quantify uncertainty about events using simulation methods
3. Assess the quality and accuracy of simulation-based descriptions of uncertainty
4. Update a description of uncertainty based on new information
5. Identify appropriate probability models for experiments/data and summarize expected outcomes from such models
6. Use correlation and conditional expectation to describe the relationship between two variables
7. Quantify uncertainty about summary statistics for large data sets

Course Details - Course Supplies

The textbook.

A calculator capable of doing simple calculations (including a square root). The calculator should not have internet connectivity. You do not need to buy a special "statistical calculator".

Course Details - Detailed Course Schedule

See above for the tentative course schedule.

Course Details - Grading Policy

Your numerical grade for this course will be computed on the basis of:

Assignment or category	Percentage
Homework	15%
Midterm 1	25%
Midterm 2	25%
Final Exam (Thursday, Dec 17, 10:00 am to 11:45 am)	35%

Your lowest homework score will be dropped before computing your grade.

Course Details - Grading Scale

Letter Grade	Percentage Range
A	100% to 93%
A-	< 93% to 90%
B+	< 90% to 87%
B	< 87% to 83%
B-	< 83% to 80%
C+	< 80% to 77%

Letter Grade	Percentage Range
C	< 77% to 73%
C-	< 73% to 70%
D+	< 70% to 67%
D	< 67% to 60%
E	< 60% to 0%

Course Details - Other

Attendance:

The course is designed around the lectures. Attendance is strongly encouraged, as active participation and consistent attendance heavily correlate with better understanding and higher course grades. While attendance is not formally graded, you will find that regular presence significantly benefits your performance.

Homework: The goal of homework assignments is to help you learn the material.

- There will be approximately eight homework assignments posted on the course website, and they will be typically due once per week, with dates and times provided. They will consist of mostly textbook-style problems and problems motivated by data analytics applications.
- Your homework solutions must be uploaded to Carmen as a **single PDF file**. Solutions can be neatly handwritten, completed on a tablet, or typed, and high-quality scanned photos are also acceptable. If you choose to handwrite your work, you are responsible for ensuring that it is clear, legible, and that the scanned version is high-quality and not blurry. All submissions must be compiled into a single PDF, with questions clearly labeled and all supporting work, including computer code, easily identifiable. Additionally, all figures and tables must be referenced and interpreted in the text. Please note that all R code and output must be typewritten to help the grader review your work efficiently.
- Late homework will not be accepted except for special circumstances,

such as illness, accidents, or other unavoidable emergencies.

- **Academic Integrity and Collaboration for Homework:** You are encouraged to work with other Stat 3201 students on homework and you may consult references both internal and external to the course material. **Each student must produce their own assignment to be handed in. Do not copy any part of another student's homework. You must list at the top of your homework your collaborators and any references (texts or other online materials) that you consulted. Use of homework solutions distributed in previous offerings of the course or available on the web constitutes academic misconduct and will be handled according to university rules.**

Exams: The goal of the exams is both to help you solidify your understanding of the material and to evaluate you on your knowledge. There will be two midterm exams and one final exam.

- All exams are in-person. More details will be posted when it comes close to the exam dates.
- All exams are closed book. You are allowed to bring a standard calculator (no smart devices) and one double-sided handwritten sheet of notes (8.5 x 11 inches).

Grade disputes: If you wish to dispute a grade on homework or an exam, you must raise the issue, in writing, within one week of the homework or exam being returned to the class. Disputes that are raised later than that will not be considered for a regrade.

Course Details - AI Use Policy

The main purpose of this course is for you to learn the material and to increase your understanding of Statistics.

- For the homework assignments, you may use AI tools as a supplement to your learning. However, you must adhere to the following requirements:
 - Write and submit your own solutions that reflect your own

understanding. Submitting AI-generated text as your own work is an act of plagiarism insofar as it involves passing off the work of others as your own.

- If you used AI to help you solve problems, below your solution, thoroughly document any use of AI: name of the AI platform, prompt you entered, and screenshot of the AI output.
- For exams, a basic calculator will be permitted, any other electronic devices will not be permitted.

Caveat: Over-reliance on AI without developing your own understanding will hinder your learning and hurt your performance on the midterms and final exam.

Responsibility for understanding: You are expected to understand your submitted homework and to be able to verbally explain to your instructor (upon request, and without advance notice) any work you submit. When you apply for jobs, you will be interviewed by professionals who instantly recognize a lack of real understanding due to over-reliance on AI; we hold you to the same standard. Do not submit what you do not understand. The response “I don’t understand what I was writing but I know it works” is never acceptable when asked to explain a submission.

Syllabus - University Policies

Office of Undergraduate Education - Standard Syllabus Statements

- [Academic Misconduct](#)
- [Religious Accommodations](#)
- [Accommodations for Disability and Illness](#)
- [Intellectual Diversity](#)
- [Grievances and Solving Problems](#)
- [Creating an Environment Free from Harassment, Discrimination, and Sexual Misconduct](#)
- [Copyrights](#)
- [Content Warnings](#)
- [Mental Health Support](#)

- [Resources for Military-Connected Students](#)
- [Campus-wide Class Cancellations](#)