

DSCI 602: Statistical Methods for Data Science

(3 Credit Hours)

Read the Course Syllabus

Learning Objectives

By the end of this reading, learners will be able to:

- State the course learning objectives
- Describe the course materials needed for the course
- Outline weekly topics
- Describe course policies
- Describe academic ethics, plagiarism policy, and other university policies

Instructor(s)

- Munni Begum, Director of Data Science and Data Analytics, Professor, Mathematical Sciences
- Jongwook Kim, Assistant Professor, Mathematical Sciences

Course Description

Data Science is a discipline at the intersection of statistics, computer science, mathematics, and an application domain. Statistical methods are at the heart of the field of data science as they involve the collection, management, processing, analysis, visualization, and interpretation of vast amounts of heterogeneous data coming from diverse disciplines. For a data scientist, it is essential to have a robust understanding of the methods used for acquiring, managing/processing, and analyzing a large volume of data.

This course equips you with the right statistical methods and tools to make sense of data effectively. We will take a systematic approach to learning about the right tools you can use. Note that as data scientists, it is important for us to be able to connect data and learn how the world around us works. To accomplish this challenging task, we will learn how

we can connect data through probability theory and statistical models and take actionable decisions, confirm a hypothesis, or make predictions.

This course provides a solid foundation in basic probability, distribution theory, methods of estimation, hypothesis testing, and basics of multivariate analysis for data science and related majors at the graduate level.

Course Objectives

After completing the course, students will be able to:

- Apply probability and distribution theory to address real-world problems related to the data science field.
- Classify the type of random variables and their probability distributions used to model various types of data in practice.
- Outline the properties of discrete and continuous random variables.
- Explain the role of normal distribution in statistical inference.
- Utilize the R computational environment for probability simulation and other statistical computing.

Suggested Study Habits

- **Read ahead.** Complete weekly reading assignments in a timely manner. Try your best not to fall behind.
- **Do not memorize.** Rather than memorizing a definition, formula, theorem, or a worked-out practical example, try to understand the basic concepts behind these. Look for their practical implications in real-life problem solving.
- **Think and analyze critically.** Look for how any given topic and lessons learned from that topic relate to other topics from the course. Always look for the big picture: how these topics fit together to give you a complete picture as in putting puzzle pieces together.

Course Materials

There is no required textbook for this course. Reading materials will be provided as needed. The following are some excellent recommended references to consult for additional reading.

Recommended Texts

- Norman Matloff. *Probability and Statistics for Data Science*, CRC Press/Taylor & Francis Group, 2020. ISBN: 978-1-136-39329-5.
<https://www.amazon.com/Probability-Statistics-Data-Science-Chapman/dp/036726093X>
- Hadley Wickham and Garrett Grolemund. *R for Data Science*, O'Reilly, 2017. ISBN: 978-1491910399.
<https://r4ds.had.co.nz/>

Software

R computational software will be used in this course to illustrate examples and problem-solving.

Course Outline

It is usually expected that students will spend approximately 2 hours of study time outside of class for every hour in class. For a 3-unit class, you should expect to study an average of 6 hours outside of class each week.

Module 1: Probability Theory – A Review

1. Probability Concepts and Definitions
2. Probability Laws
3. Conditional Probability
4. Examples
 - (a) The Bus Ridership Example
 - (b) A Simple Board Game
5. Bayes' Rule
 - (a) Bayes' Rule Examples

Module 3: Random Variables and Their Properties

1. Random Variables – Definitions and Types
2. Discrete Random Variables
 - (a) Expected Value
 - (b) Variance
 - (c) Example – Return Time for Library Books
 - (d) Example – The Bus Ridership (Expected Value Simulation)
3. Continuous Random Variables
 - (a) Expected Value
 - (b) Variance
 - (c) Examples – Expected Value and Variance of a Continuous Random Variable
4. Additional Properties of a Random Variable (Optional)
 - (a) Skewness
 - (b) Kurtosis

Module 4: Discrete Probability Distributions – Part I

1. Probability Mass Function
2. Cumulative Distribution Function
3. Bernoulli Trials
 - (a) Bernoulli Distribution
 - (b) Binomial Distribution
 - (c) Geometric Distribution
 - (d) Negative Binomial Distribution
4. Examples

Module 6: Continuous Probability Distributions – Part I

1. Continuous Probability Distributions
 - (a) Cumulative Distribution Function
 - (b) Probability Density Function
 - (c) Expected Value and Variance
2. The Uniform Distribution
3. The Normal and Standard Normal Distributions
4. Examples

Module 9: Role of Normal Distribution in Statistical Inference

1. Properties of Normal Distribution
2. The Central Limit Theorem and Normality
3. Sampling Distributions of Other Sample Statistics
 - (a) Chi-square distribution as the sampling distribution of sample variance
 - (b) Student's t distribution as the sampling distribution of the standardized sample mean
 - (c) F distribution as the sampling distribution of the proportion of two sample variances

Course Policies

Your preparedness, participation, and completion of weekly activities on time will significantly impact your course grade. Regularity and timeliness are desirable in order to complete this course successfully.

Course Assignments and Assessments

Course Assignments

Assignment details, including instructions, relevant readings, due dates, and grading criteria/rubrics, will be posted and updated on Coursera. Please check Coursera regularly to stay informed and on track.

Grading Policy

It is my policy that appropriate evaluation of your academic performance is an integral part of your learning experience. In the absence of mistake, fraud, bad faith, or incompetence, I will be the key decision-maker on the assignment of grades. For information on grade appeal, consult the Ball State website.

Grading Scale

Percentage Lower Bound	Letter Grade
93%	A
90%	A-
88%	B+
82%	B
80%	B-
78%	C+
72%	C
70%	C-
68%	D+
62%	D
60%	D-
0%	F

Course Withdrawal Statement

You can withdraw from this course within the withdrawal period. For details, see the Registrar Office's withdrawal page as well as Withdrawal Procedures in the current graduate catalog.

University and Course Policies

We are committed to ensuring that all members of the community are welcome, through valuing the various experiences and worldviews represented at Ball State and among those we serve. We promote a culture of respect and civil discourse. If you need course adaptations or accommodations because of a disability, please contact the instructor of record as soon as possible. Ball State's Disability Services Office coordinates services for students with disabilities; documentation of a disability needs to be on file in that office before any

accommodations can be provided. Disability Services can be contacted at 765-285-5293 or dsd@bsu.edu.

Freedom of Expression

In this course, we are committed to fostering a learning environment that values intellectual diversity, encourages free expression, and promotes open inquiry. As members of the Ball State Community, we treat each person with civility, courtesy, compassion, and dignity and respect and learn from differences in people, ideas, and opinions. Please review Ball State University's Statement on Freedom of Expression, the resources on Ball State's Freedom of Expression webpage, and Ball State's Beneficence Pledge.

Ball State Academic Ethics Policy

Honesty, trust, and personal responsibility are fundamental attributes of the university community. Academic dishonesty and other forms of academic misconduct threaten the foundation of an institution dedicated to the pursuit of knowledge and will not be tolerated. To maintain its credibility and reputation, and to equitably assign evaluations of scholastic and creative performance, Ball State University is committed to maintaining a climate that upholds and values the highest standards of academic integrity. Visit the VPAA's academic integrity website (Students tab) for resources on understanding academic integrity, citing sources properly, and avoiding inadvertent academic dishonesty. To learn about BSU's academic integrity expectations and students' rights, please read the University Student Academic Ethics Policy.

Course Academic Ethics and Plagiarism Policy

When completing assignments for this class, the intent is that you are learning the content, applying it to solve problems, and need to practice in order to improve. All resources that you need to complete the solutions are covered in the course, and assignments are scaffolded to help you build your understanding. When you use online solutions, AI assistants, or others to complete your work for you, you are doing yourself a disservice and not learning the material. Using these resources is considered academic dishonesty and plagiarism. **For this course, the use of online solutions and AI assistants such as Gemini or ChatGPT are prohibited.** Use of such resources violates the course Academic Ethics policy and may result in failing the course.

Academic integrity violations include giving or receiving an unfair academic advantage (cheating), presenting someone else's ideas or work as your own (plagiarism), and falsifying academic records. Unless otherwise indicated, you must work independently. Check with me if you are unsure whether something constitutes academic dishonesty. Examples of academic integrity violations include, but are not limited to:

- Using resources not authorized by the faculty member (including devices, AI tools, hidden notes, and open books)
- Using commercial study websites to find answers to graded assignments (Chegg, Course Hero, StudyPool, OneClass, etc.).
- Without the instructor's written permission, giving away, buying, or selling graded assignments, class notes, exams, study guides, or other course materials to other students or to third-party vendors (Course Hero, Chegg, 24HourAnswers, etc.).
- Working with another person on any assignment other than authorized group projects.
- Sharing or allowing others to access your files, whether done with permission or not.
- Reusing your own work from another semester, course, or section.
- Discussing exam questions and answers with students who have not taken the exam.
- Soliciting others to complete work for you, both human and AI assistants such as ChatGPT, Gemini, or other LLM-based tools.

Subject To Change Statement

This syllabus and schedule are subject to change in the event of extenuating circumstances.