CS 325 Introduction to Data Science

Winter 2024

Course Credits: 4
Contact Hours: 56 hours
Instructor: TBA
Email:TBA

COURSE OBJECTIVES

This course is designed to provide students with a comprehensive introduction to the fundamental concepts, techniques, and tools that underpin the field of data science. In an era where data is increasingly recognized as a valuable resource across various industries, this course equips students with the essential skills and knowledge necessary to harness the power of data for decision-making, analysis, and problem-solving.

Upon Completion of this Course, students will be able to:

1. Express the significance of statistics in a data-abundant environment, encompassing contemporary issues like ethical concerns, data privacy, and the vastness of big data；

2. Determine the dataset's underlying study design and assess how this design influences context-specific results；

3. Analyze the application of statistics in media and research articles across diverse data scenarios, focusing on issues related to confounding variables and bias；

4. Apply the normal approximation to data while taking into account measurement error；

5. Utilize the box model for explaining probability and probabilistic fluctuations, encompassing sample surveys and the central limit theorem.

PREREQUISITES

N/A

GRADING

Grades will be determined by accumulating points, with 100 points being the maximum, as follows:

ITEM	POINTS
3 Homework	15 Points
2 Quizzes	20 Points
Midterm Exam	20 Points
Group Project	15 Points
Final Exam	30 Points
Total	100 Points

Late submissions will be graded at the end of the course. Grades will be assigned according to the following rule:

$A \geq 90 > B \geq 80 > C \geq 70 > D \geq 60 > F$.

We reserve the right to make adjustments to the overall grading policy.

COURSE MATERIALS

Required Texts:

Jeffrey S. Saltz; Jeffrey M. Stanton, ***An Introduction to Data Science***, 1st Edition, SAGE Publications, 2018.

Recommended (Optional) Texts or Other Materials:

None

COURSE TOPICS

MODULE	TASKS
Module 1	**Topics:** Topic 1: The Steps in Doing Data Science Topic 2: Storing Data—Using Bits and Bytes Topic 3: Creating a Data Set in R Topic 4: Identifying Data Problems **Assessments:** Homework#1

	Quiz#1
Module 2	**Topics:** Topic 5: Getting Started With R Topic 6: Creating and Using Vectors Topic 7: Data Munging Topic 8: Creating Functions in R **Assessments:** Homework#2
Module 3	**Topics:** Topic 9: Sample in a Jar Topic 10: Law of Large Numbers and the Central Limit Theorem Topic 11: Importing Data Using RStudio Topic 12: Comparing SQL and R for Accessing a Data Set **Assessments:** Homework#3 Midterm Exam
Module 4	**Topics:** Topic 13: Pictures Versus Numbers Topic 14: Basic Plots in R Topic 15: Using ggplot2 Topic 16: More Advanced ggplot2 Visualizations **Assessments:** Group Project Quiz#2
Module 5	**Topics:** Topic 17: Lining Up Our Models Topic 18: Association Rules Mining Topic 19: Supervised Learning via Support Vector Machines Topic 20: The Tools for Big Data **Assessments:** Final Exam

ATTENDANCE

1) Class attendance is required. Missing classes without permission will lead to decrease in overall grade.

Missing less than two classes: no penalty.

Missing more than two classes: 7% will be taken off from the overall grade.

If the instructor reports a student's frequent missing of class to the Soochow

University Academic Administration Office, the student might get a written warning and might be prohibited from attending final exam.

2) Participants in this course are expected to arrive in class promptly and adequately prepared. The primary objective of this course is to critically engage with the readings and the subject matter. Therefore, course participants are expected to have completed the reading prior to class and prepare thoughtful reflections/commentaries to share with fellow colleagues.

LEARNING REQUIREMENTS

1) Late assignments are not acceptable and are subjected to grade deductions.
2) Assignments submitted in the wrong format will be counted as not submitted.
3) Failure to submit or fulfill any required course component results in failure of the class.
4) Make-up for midterm and final exams only with valid excuses, as defined by the University.
5) In order to earn a Certificate of Completion, participants must thoughtfully complete all assignments by stated deadlines and earn an average quiz score of 50% or greater.

TECHNOLOGY POLICY

The use of electronic devices in class is distracting, both for the user and for the rest of the class. Only non-programmable calculators can be used in the tests and exam. Any attempts to use cell phones and other electronic communication devices will be seemed as cheating. Laptops are discouraged, unless you use them for activities DIRECTLY related to the course (eg., note taking, reading course documents).

ACEDEMIC INTEGRITY POLICY

Soochow University highly values the academic integrity and aims to promote the academic fairness, honesty and responsibility. Any academic dishonesty behaviors and any attempts to cheats and plagiarism will be reported to the university administration office. A written warning and the relevant penalties will be imposed. The record might be shown on the official university transcript.

DISABILITY ACCOMMODATION

Soochow University is committed to maintaining a barrier-free environment so that students with disabilities can fully access programs, courses, services, and activities at Soochow University. Students with disabilities who require accommodations for access to and/or participation in this course are welcome.

Note:

Please contact the University Administrative Office immediately if you have a learning disability, a medical issue, or any other type of problem that prevents professors from seeing you have learned the course material.