San José State University
Department of Applied Data Science

DATA 220
Mathematical Methods for Data Analysis

Section 11
Spring 2021

Course and Contact Information

Instructor:	Ron Mak
Office location:	ENG 250 (but working from home)
Email:	ron.mak@sjsu.edu
Website:	http://www.cs.sjsu.edu/~mak/
Office hours:	TuTh 4:30 – 5:30 PM online via Zoom
Class days/time:	Th 6:00 – 8:45 PM online via Zoom
Classroom:	Zoom
Prerequisites:	Instructor consent

Course Catalog Description

Mathematical and statistical methods for data analytics. Selected topics from probability, statistics, linear algebra, and mathematical optimization. Programming for numerical implementation of mathematical and statistical algorithms.

Course Format

This course adopts a synchronous online classroom delivery format. To participate in classroom activities, submit assignments, and take tests/exams remotely, a student must have a computer with adequate internet connection and bandwidth for accessing Canvas and attending Zoom video meetings. A smartphone or tablet with a camera capable of running Zoom is also needed for video recording of your test environment during the tests/exams.

Faculty Web Page and Canvas

Course materials, syllabus, assignments, grading criteria, exams, and other information will be posted at my faculty website at http://www.cs.sjsu.edu/~mak and on the Canvas Learning Management System course login website at http://sjsu.instructure.com. You are responsible for regularly checking these websites to learn of any updates. You can find Canvas video tutorials and documentations at http://ges.sjsu.edu/canvas-students.

Course Goals

Become familiar with the mathematical foundations for data analytics, including probability, statistics, and linear algebra. Obtain practical experience solving problems with the Python data analytics modules and functions.

Course Learning Outcomes (CLO)

Upon successful completion of this course, you will:

CLO 1: Understand the probabilistic and statistical foundations of data analytics.
CLO 2: Understand and apply linear algebra for data analytics.
CLO 3: Develop programming skills to facilitate objectives 1 and 2.

Recommended Texts

You can find most of what you'll need online. Sample tutorials:

- Python: https://www.python-course.eu/python3_course.php
- Regular expressions: https://www.regular-expressions.info/tutorial.html
- Statistics and probability: https://www.khanacademy.org/math/statistics-probability
- Linear algebra: https://www.youtube.com/watch?v=kjBOesZCoqc

But if you like to read textbooks, here are some suggestions. Each of these books has many more topics in much greater depth than this course will cover.

Title: Author: Publisher: ISBN:	*Introduction to Statistics & Data Analysis*, 6th edition Roxy Peck, Tom Short, and Chris Olsen Cengage, 2019 978-1-337-79361-2
Title: Author: Publisher: ISBN:	*Linear Algebra: A Modern Introduction*, 4th edition David Poole Cengage Learning, 2015 978-1-285-46324-7
Title: Author: Publisher: ISBN:	*Linear Algebra* (Schaum's Outline), 6th edition Seymour Lipschutz and Marc Lipson McGraw-Hill, 2018 978-1-260-01144-9
Title: Author: Publisher: ISBN:	*Data Science from Scratch: First Principles with Python*, 2nd edition Joel Grus O'Reilly, 2019 978-1-492-04113-9
Title: Author: Publisher: ISBN:	*Python Data Science Handbook* Jake VanderPlas O'Reilly, 2017 978-1-491-91205-8

Software to Install

During the first class meeting, you will learn how to install the required software for the course, including Anaconda, Python, Jupyter Notebooks, and Jupyter Lab. There may be other software packages to install during the semester.

Course Requirements and Assignments

You are expected to actively participate in classroom discussions, which often lead to a deeper understanding of the concepts and are also strongly associated with your course grade.

Each class meeting will have one or more lab sessions. Since students generally learn better working together, each of you will choose a lab partner for the semester to work on the lab assignments. Each team of two students will turn in one joint set of programs and results per lab assignment, and each student on a team will receive the same score.

All the students may collaborate on assignments, but each team must write independent coding solutions. Copying and other forms of cheating will not be tolerated and will result in a zero score for the assignment (minimal penalty) or a failing grade for the course, possibly combined with other disciplinary actions from the university.

You must submit assignments on time to receive full credit. The course will end with a comprehensive final exam. No make-up exams will be given if a student misses the midterm exam (unless you have a legitimate excuse such as illness or other personal emergencies and can provide documented evidence). You will each take the midterm and final exams individually.

You must show all your work for both assignments and tests. Note that we will grade your work in terms of correctness, completeness, and clarity, not just your answer. Thus, correct answers with no or poorly written supporting steps may receive very little credit.

Grading Information

Your total score will be computed with these weights:

50% Assignments*
20% Midterm exam**
30% Final exam**

* team scores
** individual scores

Each assignment and exam will be scored (given points) but not assigned a letter grade. The average score of each assignment and exam will be available in Canvas after it has been graded.

You will be able to keep track of your total weighted score during the semester. This score determines your grade for the class at the end of the semester:

93 – 100%	A	73 – 76%	C
90 – 92%	A-	70 – 72%	C-
87 – 89%	B+	67 – 69%	D+
83 – 86%	B	63 – 66%	D
80 – 82%	B-	60 – 62%	D-
77 – 79%	C+	< 60%	F

Zoom Classroom Etiquette

- **Mute your microphone.** To help keep background noise to a minimum, make sure you mute your microphone when you are not speaking.
- **Be mindful of background noise and distractions.** Find a quiet place to "attend" class, to the greatest extent possible.
    - Avoid video setups where people may be walking behind you, people talking, making noise, etc.
    - Avoid activities that could create additional noise, such as shuffling papers, listening to music in the background, etc.
- **Position your camera properly.** Be sure your webcam is in a stable position and focused at eye level.
- **Limit your distractions and avoid multitasking.** You can make it easier to focus on the meeting by turning off notifications, closing or minimizing running apps, and putting your smartphone away (unless you are using it to access Zoom).
- **Use appropriate virtual backgrounds.** If using a virtual background, it should be appropriate and professional and should not suggest or include content that is objectively offensive and demeaning.

Recording Zoom Classes

This course or portions of this course (i.e., lectures, discussions, student presentations) will be recorded for instructional or educational purposes. The recordings will be posted to the class webpage. The recordings will be deleted at the end of the semester. **If you prefer to remain anonymous** during these recordings, then please communicate with the instructor about possible accommodations (e.g., temporarily turning off identifying information from the Zoom session, including student name and picture, prior to recording).

Students are Not Allowed to Record

Students are prohibited from recording class activities (including class lectures, office hours, advising sessions, etc.), distributing class recordings, or posting class recordings. Materials created by the instructor for the course (syllabi, lectures and lecture notes, presentations, etc.) are copyrighted by the instructor. This university policy (S12-7) is in place to protect the privacy of students in the course, as well as to maintain academic integrity through reducing the instances of cheating. Students who record, distribute, or post these materials will be referred to the Student Conduct and Ethical Development office. Unauthorized recording may violate university and state law. It is the responsibility of students that require special accommodations or assistive technology due to a disability to notify the instructor.

Proctoring Software and Exams

Exams will be proctored in this course through Respondus
Monitor, LockDown Browser, and Zoom video meeting. Please note it is the instructor's discretion to determine the method of proctoring. If cheating is suspected the proctored videos may be used for further inspection and may become part of the student's disciplinary record. Note that the proctoring software does not determine whether academic misconduct occurred but does determine whether

something irregular occurred that may require further investigation. Students are encouraged to contact the instructor if unexpected interruptions (from a parent or roommate, for example) occur during an exam. Please refer to the online exam instructions for details of the setup and requirements.

Technical Difficulties

- **Internet connection issues:** Canvas autosaves responses a few times per minute as long as there is an internet connection. If your internet connection is lost, Canvas will warn you but allow you to continue working on your exam. A brief loss of internet connection is unlikely to cause you to lose your work. However, a longer loss of connectivity or weak/unstable connection may jeopardize your exam.
- **Other technical difficulties:** Immediately notify the instructor and explain the problem you are facing. Your instructor may not be able to respond immediately or provide technical support. However, the current state of your exam and communication will provide a record of the situation.

Contact the SJSU technical support for Canvas:

Technical Support for Canvas
Email: ecampus@sjsu.edu
Phone: (408) 924-2337
https://www.sjsu.edu/ecampus/support/

University Policies

Per University Policy S16-9, university-wide policy information relevant to all courses, such as academic integrity, accommodations, etc. will be available on Office of Graduate and Undergraduate Program's Syllabus Information web page at http://www.sjsu.edu/gup/syllabusinfo/.

DATA 220
Mathematical Methods for Data Analysis

Spring 2021

Course Schedule (subject to change with fair notice)

Week	Date	Topics
1	Jan 28	**Introduction to Data Analytics** What is data science? **Lab:** Install Anaconda **Lab:** Data acquisition, CSV files → **`pandas`** dataframes
2	Feb 4	**Descriptive Statistics** Python lists, tuples, and dictionaries **`numpy`** arrays Random values **Lab:** Histograms with **`matplotlib`** and **`seaborn`**
3	Feb 11	Median, quartiles and percentiles Measures of central tendency Measures of dispersion Static data visualizations with **`matplotlib`** and **`seaborn`** **Lab:** Basic descriptive statistics
4	Feb 18	**Probability** Counting principles Discrete and continuous random variables Uniform, normal, exponential, binomial, and Poisson distributions **Lab:** Dynamic data visualizations with **`matplotlib`** animation
5	Feb 25	Expected value Order and repetition Independent and dependent events Conditional probability Bayes' Theorem **Lab:** Probability problem set
6	Mar 4	Tactics for solving probability problems The Monty Hall Problem **Statistical Analysis** Population vs. sample statistics Random sampling **Lab:** The Central Limit Theorem
7	Mar 11	Hypothesis testing Confidence intervals Levels of significance Type I and Type II errors

Week	Date	Topics
8	Mar 18	***Midterm exam*** Small sample hypothesis tests Testing two population means
9	Mar 25	P-values Correlation Scatter plots Linear regression with least squares fit Logistic regression **Lab:** Linear regression analysis
	Mar 29 – Apr 2	***Spring break***
10	Apr 8	**Introduction to Statistical Applications in Data Analytics** Time series analysis and forecasting Supervised and unsupervised learning Training and test data Introduction to machine learning (ML) Time series analysis with ML Multiple linear regression with ML Training models for supervised ML Underfitting and overfitting **Lab:** Multiple regression analysis
11	Apr 15	Classification algorithms Unsupervised machine learning Dimensionality reduction Principal component analysis Natural language processing with Python **Lab:** k-nearest neighbor classification and k-means clustering
12	Apr 22	**Linear Algebra** Vectors and matrices Systems of linear equations Augmented matrices and row echelon form Gaussian elimination Matrix inverse **Lab:** Statistics problem set
13	Apr 29	System of linear equations Augmented matrix Matrix form of the least squares solution QR factorization LU decomposition Singular value decomposition (SVD) **Lab:** Matrix operations
14	May 6	Nonlinear relationships Polynomial regression Markov chains and steady state **Lab:** Polynomial regression and Markov chains

Week	Date	Topics
15	May 13	Eigenvalues and eigenvectors Geometric interpretations
Final Exam	***Thursday May 20***	Time: 5:15 – 7:30 PM Zoom