

36-202 (Spring 2026)

Methods for Statistics and Data Science

Location and Times:

Spring 2026
Sunday & Tuesday 8:30 - 9:45 AM
Classroom 3035

Instructor Information:

Taeyong Park, Ph.D.
✉taeyongp@andrew.cmu.edu
Office: 2191
Zoom Office: <https://cmu.zoom.us/my/tpark>
Office Hours: Sunday 10:00 - 11:00 AM & Monday 2:30 - 3:30 PM
TA: Wijdan Tariq (✉twijdan@andrew.cmu.edu)
CA: Razan Murshid (✉rmurshid@andrew.cmu.edu)

1. Course Description

This is the second course in the sequence of statistical data analysis at CMUQ. Building on probability theory, core concepts in statistics, and basic computer skills covered in 70-207 or 36-200, you are expected to learn more advanced statistical methods and computational skills for rigorous statistical analysis. You learn mathematical foundations for linear regression analysis and the ordinary least squares estimator (OLS). In addition, logistic regression, multinomial logistic regression, and ordinal logistic regression are taught, which lead to comprehensive discussions about how to build statistical models for different types of response data. Finally, the machine learning perspective and basic classification and clustering algorithms are discussed. Throughout the semester, you use statistical environment R (<http://www.r-project.org/>) which is widely used in business, economics, biostatistics, social science, and many other fields. You use R to manage and analyze data. Hands-on applications of statistical inference to questions in your field of study are highly emphasized. You are expected to come up with your own research question and conduct data analysis for the research project assignment.

2. Prerequisites

- 36-200 or 70-207.

3. Textbooks

- David R. Anderson et al. 2017. *Essentials of Statistics for Business and Economics*. 8th Edition. CENGAGE Learning.

4. Requirements and Evaluation

Attendance, Quiz, and Participation (4%): You are required to attend every class on time and remain until the end. However, I understand that students may occasionally need to miss class for reasons such as illness or emergencies. Your **first three absences will not affect your final grade**. Beginning with the 4th absence, each additional absence will result in a deduction of 0.5% from your final grade. **Please note: I do not distinguish between excused and unexcused absences, and you do not need to provide medical documentation.** Attendance will be taken by the TA and CA. If you arrive more than 15 minutes late, you will be counted as absent. If you leave class for more than 10 minutes or do not return, you will also be counted as absent. To ensure accurate tracking of attendance and participation, we will use designated seating. For any questions about your attendance record, please contact the CA, TA, or me.

In addition, up to 2% bonus points will be awarded for **exceptionally** active class participation based on the two criteria below.

- On randomly selected days, you will be given a short in-class quiz consisting of 2–3 simple questions and submit your answers during class. You must submit all quizzes throughout the semester.
- Active participation includes, but is not limited to, asking relevant questions, answering my questions, volunteering to share thoughts, etc. Active participation throughout the semester is determined at my discretion.

Exercise and Recitation (10%): Exercise questions are assigned for each module throughout the semester as a homework assignment. The submission deadlines are in the course outline below, and you should submit your assignment to TA Wijdan Tariq.

- The exercise is designed to motivate you to keep up with the material on a regular basis. Furthermore, it provides you with practice questions for the exams.
- Your grade for each exercise will be 1 for complete; 0.5 for submitted but incomplete; 0 for no submission. If you show sufficient effort on your submitted exercise, it will be considered complete, regardless of correct/incorrect answers. Sufficient effort requires evidence that you have worked on at least 80% of the entire questions.
- The answer key will be posted on Canvas after the submission deadline, and hence no extension is permitted. You are responsible to check that your work is correct using the answer key.

Mid-term Exam (17%): An in-class exam covering all material discussed up to the exam date.

Research Project (34%): A group project focused on linear and logistic regression analysis.

Final Exam (35%): In-class exam during the final exam week.

Letter Grade Distribution:

>= 90.00	A (Exceptional)
80.00 - 89.99	B (Good)
70.00 - 79.99	C (Satisfactory)
60.00 - 69.99	D (less than satisfactory, but passing)
<= 59.99	R (Failing to pass)

5. Office Hours and Appointments

I hold office hours: Sunday 10:00 - 11:00 AM & Monday 2:30 - 3:30 PM. I encourage you to set up an appointment to make sure I will be available if you want to see me other than during my office hours.

6. Academic Integrity

You must comply with the academic integrity policy. You are required to refer to CMU's general policies on cheating and plagiarism: <http://www.cmu.edu/academic-integrity/valuing/index.html>. Violations of CMU's general policies on cheating and plagiarism carry a range of consequences: <http://www.cmu.edu/academic-integrity/understanding/index.html>.

You may use generative AI programs like ChatGPT during the brainstorming and idea generation phase for assignments. However, doing so cannot be considered a substitute for traditional research. Generative AI programs rely on predictive models to generate content that may appear correct, but has been shown to sometimes be incomplete, inaccurate, taken without attribution from other sources, and / or biased. Any information generated by an AI program should be cited like any other reference material. You are ultimately responsible for the content of the information you submit. However, you may not attempt to pass off any work generated by an AI program as your own.

7. Disability Resources and Health and Well-being

You can find information about disability-related accommodations on <https://scotty.qatar.cmu.edu/health-and-wellness/medical-accommodations/>. You may also consult me or CMUQ staff (Office of Health and Wellness) regarding learning disabilities, health, and wellness.

8. Diversity, Equity, and Inclusion

It is critical for me to ensure that students from all diverse backgrounds and perspectives feel belonging to this course, that students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally or for other students or student groups. In addition, if anything conflicts with your value based on gender, sexuality, disability, age, socioeconomic status, ethnicity, race, and culture, please let me know so that I can make arrangements for you.

9. Course Outline

Date	Topic	Reading
INTRODUCTION		Syllabus
Jan 6	Course Introduction	
MODULE 1: Bivariate Linear Regression		Ch.14.1–14.6
Jan 11, 13	Lecture, Lab, & Exercise	
Jan 15	Recitation	
Homework	Exercise 1 Due by Jan 14 11:59 PM	
MODULE 2: Multivariate Linear Regression		Ch.15.1–15.6
Jan 18, 20	Lecture, Lab, & Exercise	
Jan 22	Recitation	
Homework	Exercise 2 Due by Jan 21 11:59 PM	
MODULE 3: Residuals, R^2, and F Test		Ch.15.1–15.6

Date	Topic	Reading
Jan 25, 27	Lecture, Lab, & Exercise	
Jan 29	Recitation	
Homework	Exercise 3 Due by Jan 28 11:59 PM	
Feb 1	Review for Exam 1	
Feb 3	Exam 1	
MODULE 4: Regression Diagnostics and Transformations		
Feb 8, 15, 17	Lecture, Lab, & Exercise	
Feb 19	Recitation	
Homework	Exercise 4 Due by Feb 18 11:59 PM	
Feb 23, 25	Spring Break; No Classes	
MODULE 5: Categorical Independent Variables		Ch.15.7
Mar 1, 3	Lecture, Lab, & Exercise	
Mar 5	Recitation	
Homework	Exercise 5 Due by Mar 7 11:59 PM	
MODULE 6: Interactions		
Mar 8, 10	Lecture, Lab, & Exercise	
Mar 12	Recitation	
Homework	Exercise 6 Due by Mar 14 11:59 PM	
MODULE 7: Regression Analysis with Discrete Dependent Variables I		Ch.15.9
Mar 15, 17	Lecture, Lab, & Exercise	
Mar 19	Recitation	

Date	Topic	Reading
Homework	Exercise 7 Due by Mar 30 11:59 PM	
Mar 22, 24	Eid Break; No Classes	
MODULE 8: Regression Analysis with Discrete Dependent Variables II		Ch.15.9
Mar 29, 31, Apr 5	Lecture, Lab, & Exercise	
Apr 9	Recitation	
Homework	Exercise 8 Due by Apr 11 11:59 PM	
Apr 6	Paper Submission Due by 11:59 PM	
Apr 7	Presentation	
Apr 12	Oral Test	
MODULE 9: Classification Methods and Clustering Methods		
Apr 14, 19, 21	Lecture, Lab, & Exercise	
Apr 23	Recitation	
Homework	Exercise 9 Due by Apr 25 11:59 PM	