70-208 (Spring 2025)

Regression Analysis

Location and Times:
Spring 2025
Monday & Wednesday 2:30 - 3:45 PM
Classroom 1185

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 & Wednesday 11:30 AM - 12:30 PM
TA: Ishaq Ansari (✉iansari@andrew.cmu.edu)
CA: Boushra Bendou (✉bbendou@andrew.cmu.edu)
CA: Aaliya Savira (✉dsavira@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 statistical computing skills covered in 70-207 or 36-200, you are expected to learn more advanced topics for rigorous statistical analysis. We use regressions to summarize the relationship in the sample between a dependent variable and one or more independent variables. We can use the relationship in the sample to estimate the relationship in the population and conduct an inference about it. This allows us to determine the precision of our estimates and assess the accuracy of the predictions we make. To conduct this inference, we use statistical theory. Throughout the course, we will look at a number of applications to show how these methods are used. This will involve both in-lass examples and assignments that give you the opportunity to work with the data on your own. Lectures and assignments will teach how to use these methodologies and interpret the results, allowing you to develop the skills to be informed users of regression methodologies. We will do a number of economics and business-related applications to highlight how these tools will be useful in the workforce.

2. Learning Objectives

By the end of the course, you will be able to do the following:

- Interpret the coefficients and determine the accuracy of coefficient estimates and predicted values from various regression analyses, including linear, logistic, multinomial logistic, and ordinal

logistic.

- Judge the goodness of fit of a regression.
- Write down the regression model best suited to answer the question of interest.
- Test hypotheses involving parameters of the regression model.
- Use R to analyze data to conduct regression analysis.

3. Prerequisites

- 70-207 or 36-200.

4. Textbook and Reading Materials

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

5. Requirements and Evaluation

Exercise and Recitation: Excercise questions are provided for each module. You should solve them and bring questions to the recitation. No grade on the exercise questions.

Problem Sets (8%): Eight problem sets will be assigned throughout the semester as a homework assignment. The submission deadlines are in the course outline below. Please submit a paper copy of your problem set to the Ishaq Ansari .

- The problem set 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. Each problem set is worth 1% of the final grade, leading to 8% as total.
- Your grade for each problem set will be 1 for complete; 0.5 for submitted but incomplete; 0 for no submission. If you show sufficient effort on the paper copy of your problem set, 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. No extension is permitted. You are responsible to check that your work is correct using the answer key.

Participation and Quiz (6%): In each class, you will be given a few (2-4) simple questions. You will turn in your answers during class. Your grade for each day will be 1 for submitted and 0 for not submitted. Your **four lowest grades** will be dropped at the end of the semester, including "0" grades for no submission due to absence. **PLEASE NOTE THAT I do not distinguish between excused and unexcused absences.** Regardless of the reason, a missed submission will count as a "0". You do not need to provide medical documentation for your absence.

In addition, I will award bonus points (up to 2% of the final grade) for exceptionally active class participation **throughout the semester**, such as asking relevant questions, answering my questions, volunteering to share thoughts, etc. These bonus points will be based on my subjective judgment.

Exam 1 (17%): In-class exam. This covers the material discussed until the exam day.

Exam 2 (17%): In-class exam. This covers the material discussed after Exam 1 and until the exam day.

Exam 3 (17%): In-class exam. This covers the material discussed after Exam 2 and until the exam day.

Final Research Project and Presentation (35%): You must submit a research paper by the deadline. You will present your paper on the final exam day.

- The paper must consist of **two hypotheses** that will be tested by 1) linear regression analysis and 2) one of logistic, multinomial logistic, or ordinal logistic regression analysis. The length of the paper must be between 5,000 and 6,000 words (excluding a cover page, tables, figures, and references); 12 points, double space, and 1-inch margins all around.
- This is a team project. You should form a team of three or four. **Late submission** will be allowed, with a penalty per day. For instance, if you submit your assignment within 24 hours after the deadline, the full credit you possibly get would be 96 points, instead of 100 points, if you submit your assignment between 24 and 48 hours after the deadline, the full credit you possibly get would be 92 points, and so on and forth.
- For your research topic, you can choose your own research question and collect a dataset for the question. I encourage you to choose a case that is relevant to your domain knowledge, i.e. business or economics. Alternatively, I will provide several data cases for you to choose from. You will need to submit your **data file and R code** along with your paper by the deadline so that I can replicate your analysis.
- **Presentation**: Submit your presentation material (for example, ppt) and present it on the final

exam day. You will have about 10 minutes for the presentation and 2 minutes for Q & A. You are also required to listen to others' presentations and ask relevant questions.

- Further details with a grading rubric will be provided later in the semester.

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)

6. Electronic Devices and Punctuality

I expect you to be respectful to me and your fellow students to create an environment that is most conducive to learning.

- You will often use your laptop or desktop during class this semester. However, this does not mean that you can feel free to use the computer for whatever you want. It is important to **use it only for class purposes** so that you will not distract yourself and you will not disrupt your classmates. Furthermore, **your cell phone must be turned off** during class. If there is an emergency that might oblige you to be contacted, please talk to me before class. I quote the following passage from the Qatar Business Administration Program Classroom Conduct, which, I believe, must apply to other programs as well:

  - *Laptops are to be closed. When class is in session, you may use your laptop only as directed by your professor. You should not check email, tweet, text, play games, or surf the Internet, any activity that diminishes your or your classmates' engagement with the classroom content and process. If you are unsure whether a given activity is appropriate, ask your professor. This policy extends to all electronic devices. Be sure that your phones and tablets are silenced and stowed before the class begins. Professors may add specific limits on the sharing or use of personal electronics in exam situations.*

- **You must come to class on time and remain in class** once the class has begun. I quote the following passage from the Qatar Business Administration Program Classroom Conduct, which, I believe, must apply to other programs as well:

- *In common business culture, punctuality is an important part of showing respect for your colleagues and business partners. Showing up late for a meeting tells the others involved that you do not place much value on their time. QBA students will demonstrate respect for their courses, classmates and professors by arriving for class early enough to get settled and prepared before the scheduled meeting time.*

7. Office Hours and Appointments

I hold office hours: Sunday 10:00 - 11:00 AM & Wednesday 11:30 AM - 12:30 PM or by appointment. 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.

8. 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.

9. 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.

10. 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.

11. Course Topics

1. Review of EDA and inference about the population mean
    - Population vs. sample
    - Joint, marginal, conditional probabilities
    - Confidence intervals and hypothesis testing
    - Covariance and correlation coefficient
    - Barplot, boxplot, histogram, scatterplot
2. Bivariate linear regression and OLS
    - Least squares line
    - Least squares algebraic properties
    - Assessing the fit of a regression
    - Assumptions
    - Confidence intervals and hypothesis testing
    - Predictions
3. Multivariate linear regression and OLS
    - Correlation vs. causality
    - Omitted variable bias
    - Interpretation of results
    - Confidence intervals and hypothesis testing
    - R^2, $Adjusted\ R^2$, F-tests
    - Predictions

4. Functional forms and transformation
    - Polynomials
    - Transformation of the dependent and/or the independent variables
5. Categorical independent variables and interaction
    - Dummy variables
    - Fixed effects
    - Interaction effects
6. Binary dependent variables
    - Logistic regression
    - Interpretation of results and significance tests
    - Predictions
7. Nominal and ordinal dependent variables
    - Multinomial logistic regression
    - Ordinal logistic regression
    - Interpretation of results and significance tests
    - Predictions
8. Applications to Business Case Studies
    - Regression results to decision making

12. Course Outline (subject to change)

Date	Topic	Reading
INTRODUCTION		Syllabus
Jan 6	Course Introduction	
MODULE 1: Review of Statistical Tests and Bivariate Linear Regression		Ch.14.1-14.6
Jan 8, 13, 15	Lecture & Exercise	
TBD	Recitation	
Homework	Problem Set 1 Due by Jan 19 5:00 PM	
MODULE 2: Multivariate Linear Regression and Model Fit		Ch.15.1-15.6
Jan 20, 22, 27	Lecture, Lab, & Exercise	
Jan 29	Short break; No Class	
TBD	Recitation	
Homework	Problem Set 2 Due by Feb 2 5:00 PM	
Feb 3	Class Activity: Applications to Business Cases	
Feb 5	**Exam 1**	
MODULE 3: Regression Diagnostics and Transformations		Ch.15.1-15.6
Feb 10, 12	Lecture, Lab, & Exercise	
TBD	Recitation	
Homework	Problem Set 3 Due by Feb 16 5:00 PM	
MODULE 4: Linear-log, Log-linear, and Log-log Models		
Feb 17, 19	Lecture, Lab, & Exercise	
TBD	Recitation	
Homework	Problem Set 4 Due by Mar 2 5:00 PM	
Feb 24, 26	Spring Break; No Classes	
MODULE 5: Categorical Independent Variables		Ch.15.7
Mar 3, 5, 10	Lecture, Lab, & Exercise	
TBD	Recitation	
Homework	Problem Set 5 Due by Mar 9 5:00 PM	
Mar 12	**Exam 2**	

Date	Topic	Reading
MODULE 6: Interactions		
Mar 17, 19	Lecture, Lab, & Exercise	
TBD	Recitation	
Homework	Problem Set 6 Due by Mar 23 5:00 PM	
MODULE 7: Regression Analysis with Discrete Dependent Variables I		Ch.15.9
Mar 24, 26	Lecture, Lab, & Exercise	
TBD	Recitation	
Homework	Problem Set 7 Due by Apr 7 5:00 PM	
Mar 31, Apr 2	Eid al-Fitr; No Classes	
MODULE 8: Regression Analysis with Discrete Dependent Variables II		
Apr 7, 9, 14	Lecture, Lab, & Exercise	
TBD	Recitation	
Homework	Problem Set 8 Due by Apr 15 5:00 PM	
Apr 16	**Exam 3**	
Apr 20	Class Activity: Applications to Business Cases	
Apr 22	Workshop for the Final Project	