Last Updated 09/07/2021

Master of Science in Data Analytics/Binghamton University

DATA 501: Predictive & Inferential Analytics

Fall 2021 Syllabus

Instructor Information

Xingye Qiao
qiao@math.binghamton.edu

Office: WH-134
Research Interests: Statistics, Machine Learning, High-dimensional Inference, Big Data Analytics, Precision Medicine
Office hours: Monday 10:30 to 11:30 on Zoom or in-person. Email in advance.

Communication Policy

Preferably all communication will be done on Piazza. General questions about the course contents can be asked so that both the questions and the answers can be visible to all. Private messages can also be sent on Piazza. Questions may be posted anonymously to the other students. Please follow the Netiquette rules shown below in the syllabus when you communicate on Piazza.

Course Information

Description

Data-driven and evidence-based decision-making have become increasingly important to professionals, practitioners, and policymakers across various industries, government, and public sectors. In this course, students will study multiple linear regression, a powerful modeling tool for data analytics commonly used to quantify the relationship between predictor variables and a continuous response variable. In addition to making predictions, the emphasis of the course is on such inferential thinking as to whether and how a response variable relates to one or more predictor variables. Students will learn the fundamental theory and principles behind linear regression and other related methods for prediction and inference, and how to properly apply these methods to real-world problems using statistical software and draw valid conclusions.

Course Delivery Mode

This course will be a traditional in-person course.

Last Updated 09/07/2021

Course Objectives

- Obtain proficient programming skills and knowledge of statistical software.
- Estimate parameters and predict response values using linear regression models.
- Understand basic principles of regression models in terms of fitting, inference, and prediction, and use them to draw valid conclusions.
- Evaluate the goodness of fit for linear regression models.
- Understand the assumptions that need to be met for a valid linear regression model.
- Use diagnostic tools to check common assumptions including normality, constant variance, independence, etc.
- Know how to incorporate qualitative predictors into the regression model.
- Understand the needs for data transformation and use proper data transformation to prepare the data for more efficient data analysis and valid inference.
- Diagnose and mitigate other challenges such as collinearity, non-linearity, etc.
- Use principal component regression and ridge regression to deal with collinear data.
- Conduct model assessment and selection using selection criteria or cross-validation.
- Understand and use lasso regression models in the high-dimensional setting.
- Criticize data snooping and unethical use of regression to generate misleading results.
- Understand and overcome the difficulty of valid inference post variable selection.
- Use logistic regression, Poisson regression, and non-linear regression (optional).
- Understand the counterfactual framework of causal inference and use a variety of approaches (optional).
- Apply various regression models to analyze real-world data in an organized way.
- Design analysis pipeline to conduct regression-based data analysis.
- Communicate effectively the procedures and results in both written and verbal forms.

Prerequisite(s)

Data 500. In addition, as a minimum, the student should have taken introductory statistics, have working knowledge of one programming language, and have some experience in analyzing and drawing conclusions from data. The recommended mathematics background includes Calculus I and II and Matrix Algebra, but they are not required. Necessary statistical concepts include random variables, probability distributions, independence, mean and variance, conditional distribution and conditional expectation, estimation, confidence intervals, and hypothesis tests. Most materials have been reviewed in Data 500.

Relationship to Other Courses

This course follows Data 500: Introduction to Data Analytics. It is one of the second courses in the Master of Science in Data Analytics program and runs in parallel with Data 502. This course focuses on the regression problem with continuous response variables while Data 502 focuses on other learning tasks (such as classification with a categorical response, clustering with no response, and others). This course also has an emphasis on inference while Data 502 places more emphasis on prediction and pattern recognition.

Last Updated 09/07/2021

Topic Outline/Schedule

Lecture	Date	Topics
1	Mon - Sep 13	Introduction
2	Wed - Sep 15	Simple linear regression (1)
3	Mon - Sep 20	Simple linear regression (2)
4	Wed - Sep 22	Diagnostics for simple linear regression.
5	Mon - Sep 27	Multiple linear regression.
6	Wed - Sep 29	Multiple linear regression.
7	Mon - Oct 04	Multiple linear regression.
8	Wed - Oct 06	Diagnostics for multiple linear regression.
9	Mon - Oct 11	Diagnostics for multiple linear regression.
10	Wed - Oct 13	Interactions and qualitative variables.
11	Mon - Oct 18	Interactions and qualitative variables.
12	Wed - Oct 20	Analysis of variance.
13	Mon - Oct 25	Transformations
14	Wed - Oct 27	Weighted Least Squares
15	Mon - Nov 01	Correlated errors.
16	Wed - Nov 03	Collinear data
17	Mon - Nov 08	PCR & Ridge
18	Wed - Nov 10	Selection.
19	Mon - Nov 15	Shrinkage Estimator
20	Wed - Nov 17	Lasso and Ridge Regression
21	Mon - Nov 22	Post-selection Inference and Data Snooping
	Wed - Nov 24	BREAK - NO CLASS
22	Mon - Nov 29	Logistic regression.
23	Wed - Dec 01	Poisson regression.

24	Mon - Dec 06	Causal Inference
25	Wed - Dec 08	Causal Inference

Course Requirements

Required Text

- (**CH**) Chatterjee, S., & Hadi, A. S. (2015). *Regression Analysis By Example, 5th edition.* John Wiley & Sons.
    - Binghamton University students have free unlimited access to this book online via the libraries. Check the website of Binghamton Libraries. Search the book after logging in.
    - Download the data sets in text formats or as an R workspace.
    - Errata can be found here.

Companion Books

- (**ISLR**) James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. New York: springer. An excellent introduction to statistical learning. In terms of regression methods, it goes beyond linear models. We will use this book in the later part of the course for topics on variable selection, penalized regression, and logistic regression. A free copy is available on the authors' website: http://faculty.marshall.usc.edu/gareth-james/ISL/
- (**LMR**) Faraway, J. (2014), *Linear Models with R, 2nd edition*. Chapman & Hall/CRC Texts in Statistical Science. It is about practical linear regression and is heavy on application and coding. Cuts right to the point. Ideal for students who want more code.
- (**ALR4**) Weisberg, S. (2013). *Applied Linear Regression, 4th edition.* John Wiley & Sons. Despite "applied" in the title, this book covers more mathematical derivations. Suitable for students with backgrounds in mathematics and statistics and wish to see more mathematical foundation for linear model theory. Binghamton University students have free unlimited access to this book online via the libraries. Check the website of Binghamton Libraries.
- (**MARR**) Sheather, S. (2009), *A Modern Approach to Regression with R*. Springer Texts in Statistics. Interesting and entertaining examples. Assumes a background in linear algebra and R.

Other Readings

- Blitzstein, J. K., & Hwang, J. (2019), *Introduction to Probability*, 2nd edition. CRC Press. Covers probability theory. This book assumes a background in Calculus 3. The chapter on conditional expectation is a must-read. See corresponding lecture videos here: http://stat110.net

Other Requirements

- Internet service: Brightspace, Piazza (Privacy Policy), GradeScope (Privacy Policy), Google G Suite (including Google Colab).
- Hardware: Laptop or desktop computer
- Software:
    - R and RStudio

GradeScope and Piazza

- We will use GradeScope (https://www.gradescope.com/courses/307870) to collect homework assignments, grade them, and provide feedback.
- We will use Piazza (https://piazza.com/binghamton/fall2021/data501/home) for all the discussions and Q & A.

COVID-19 related policy

Binghamton University follows the recommendations of public health experts to protect the health of students, faculty, staff, and the community at large. Safeguarding public health depends on each of us strictly following requirements as they are instituted and for as long as they remain in force. Health and safety standards will be enforced in this course.

Face coverings and other safety measures

Current rules require **everyone to wear a face-covering that completely covers both the nose and mouth while indoors** (unless they are eating or alone in a private space like an office). **A face shield is not an acceptable substitute.** Classroom safety requirements will continue to be based on guidance from public health authorities and will be uniformly applied across campus. If these requirements change, a campus-wide announcement will be made to inform the University.

- **If you forget your face covering or it does not meet these requirements, you will be asked to leave the room immediately.** You may not return until you meet the requirement.
- If you miss a graded assessment due to being asked to leave the classroom for not having a proper face covering, **you will receive a zero on the assessment**.
- If you do not meet the face-covering requirement and refuse to leave the room when directed, the instructor will immediately cancel the remainder of the class session and inform the dean's office, which will work with the Student Records office to **issue a failing grade ("F") for the course** regardless of when in the semester the incident occurs. The dean's office will also inform the Office of Student Conduct.
- If a student's refusal to comply is a second offense, the Office of Student Conduct may recommend **dismissal from the University**.

If the rules for health and safety measures change, the campus will be notified and the new requirements will take effect.

The University recommends and supports swift action and clear consequences if a student's non-compliance risks the safety of others. The academic and course-removal sanctions listed here are provided because the Provost's Office considers them to be valid responses if a student puts the safety of others at risk. Non-compliance with safety requirements constitutes a public health risk and a disruption of the learning experience.

Illness/quarantine

Students should stay home if they are ill and seek prompt medical evaluation if they experience symptoms of COVID-19. Early case findings will benefit the entire campus. If you have to be absent due to illness or quarantine, please notify the instructor team, who will work with you to identify solutions for you to gain access to the learning materials.

Credit Hours and Expectations

This course is a 3-credit course over a 13-week period, which means that students are expected to do at least 10 hours of course-related work each week during the class. This includes 3 hours for in-class lectures, and 7 hours completing assigned readings, studying for tests and examinations, preparing written assignments, and other course-related tasks.

Assignments and Grading

The following assignments are required for the course.

- Reading assignments
- Homework assignments
- Midterm exam I
- Midterm exam II
- Group project

Final Grade Breakdown

Assignment Name	Percent of Total
Homework	50%
Midterm Exam I	15%
Midterm Exam II	15%
Group Project	18%

Participation	2%

Self-check quizzes

Self-check quizzes should ideally be done after the readings and before the lectures. They are meant to check if the student has done the reading. Quizzes are graded automatically. There is no limit on the number of attempts a student can make for each quiz. The grade for quizzes does not enter the total grade breakdown, but a student must pass all the self-check quizzes to receive a grade in the course.

Homework Assignments

The weights for the homework assignments are proportional to the number of problems. Together the homework assignments are 50% of the final grade.

Midterm Exams

There will be two midterm exams. One will be scheduled in early November and the other in early December. Each midterm exam will count for 15% of the final grade.

Group Project

This project is a comprehensive analysis of a moderately large dataset using the tools we have learned throughout the course. The project should be done by a group of no more than four students. The instructor will provide a few candidate datasets for the groups to choose from, or they can use their own dataset after seeking approval from the instructor. There are no right or wrong answers to many questions involved in the project. The goal of the project is to try to mimic the analysis of a real data set that students might come across. As a final deliverable of the project, a final report should be sent to the instructor.

Other Participation

- The rest 2% will be rewarded to students who actively participate in lectures, answer questions, actively participate in the online discussion, including asking and answering questions on Piazza.

Grading Scheme

Grade	***Percent***
A	93% – 100%
A-	90% – <93%

B+	87% – <90%
B	83% – <87%
B-	80% – <83%
C+	77% – <80%
C	73% – <77%
C-	70% – <73%
D	60% – <70%
F	0% – <60%

Accessing Grades

Students can assess their grades using Brightspace.

Extra Credit (5 points maximum)

This course was offered for the first time in Fall 2020. Your feedback will make this course better for you and future students. Give 5+ pieces of feedback/corrections to course material.

- typos/spelling/grammar (there are likely many)
- rewording awkward/unclear language
- improve visualizations
- point out explanations you found effective/could be better

Send the feedback (whenever it becomes available) to the course instructors publicly on Piazza (using the *feedback* tag; you can hide your identity, but the content should be visible to all). We will mark the feedback down and make adjustments to your grade in the end if you contribute 5 pieces or more.

Please keep an eye on the feedback as sometimes it helps to correct an error or clarify a muddy point. If similar feedback has been made, then duplicate feedback will not be counted for the extra credit purpose.

Course Policies

Penalties for Late Work and Requests for Extensions

- Late submissions are allowed for up to one week. For each day the submission is late, the grade will be reduced by 10%. For example, homework that is 3 days late will only receive 70% of the real grade should you submit the work on time.
- Students may request an extension of the work before the due time.

- Self-check quizzes will be closed one month after the teaching of the corresponding contents is over. Note that these quizzes are supposed to be finished when or before each lecture is taught. A student who did not pass a self-check quiz will receive F in the course.

Make-up Exam Policy

Students who cannot take their exams as scheduled because of religious conflict, documented serious illness, or compelling, unexpected circumstances may appeal for a make-up exam. The following are examples that are not considered compelling reasons to grant a make-up exam: lack of preparation, negligence, misinformation, or planned vacations and other events.

Absences Due to Religious Holidays

If you anticipate being absent because of any religious observance, please notify me in writing at least one week in advance. We will work together to accommodate.

Attendance & Participation

Class attendance and participation points are given to encourage your active class participation and discussion. You will be rewarded with a perfect score as long as you frequently come to class and actively contribute to the class discussion during recitations and lectures.

Although it is not required, most students send their professors a brief e-mail to explain their absence in advance. Students who repeatedly arrive late to the lecture or recitation will have their Class Participation grade lowered. Please sign the attendance sheet (or e-attendance sheet) when you come to the class. Any false signatures will result in zero participation grades for all parties involved.

Class participation is a very important part of the learning process in this course. You will be evaluated on the *quality* of your contributions and insights. Quality comments possess one or more of the following properties:

- Offers a different and unique, but relevant, perspective;
- Contributes to moving the discussion and analysis forward;
- Builds on other comments;
- Transcends the “I feel” syndrome. That is, it includes some evidence, argumentation, or recognition of inherent trade-offs. In other words, the comment demonstrates some reflective thinking.

We will use our assessment of your participation to manage borderline grades. While your participation grade is subjective, it will not be random or arbitrary. And, clearly, more frequent quality comments are better than less frequent quality comments.

Activities on the online forum are an important part of the participation points.

Incomplete

- *It is possible to enter an incomplete contract with the instructors if you cannot complete the course in time. Consult the instructor.*

Academic Integrity

Binghamton University provides explicit guidelines in the Student Academic Honesty Code (see the University Bulletin - Academic Policies and Procedures for All Students). Unless specified otherwise in the syllabus, I expect the work you submit for grading to be yours and yours alone. Not acknowledging another's work with proper references, taking credit for someone else's work, letting your work appear in another student's paper, or fabricating "results" are grounds for failing the assignment and/or the course. If you have any questions about what constitutes plagiarism or cheating, please ask me.

Netiquette

Unfortunately, there are instances where disrespectful and inappropriate behavior can occur in a course, even in an online setting. It is important to remember to be respectful of your classmates and their ideas. Here are the Netiquette Rules students are expected to follow in this course. Based on “15 Rules of Netiquette for Online Discussion Boards” by Touro’s College Online Education Department.

1. Before posting your question to a discussion board, check if anyone has asked it already and received a reply. Just as you wouldn’t repeat a topic of discussion right after it happened in real life, don’t do that in discussion boards either.
2. Stay on topic – Don’t post irrelevant links, comments, thoughts, or pictures.
3. Don’t type in ALL CAPS! If you do, it will look like you’re screaming.
4. Don’t write anything that sounds angry or sarcastic, even as a joke, because without hearing your tone of voice, your peers might not realize you’re joking.
5. Always remember to say “Please” and “Thank you” when soliciting help from your classmates.
6. Respect the opinions of your classmates. If you feel the need to disagree, do so respectfully and acknowledge the valid points in your classmate’s argument. Acknowledge that others are entitled to have their own perspective on the issue.
7. If you reply to a question from a classmate, make sure your answer is accurate! If you’re not 100% sure when the paper is due, DO NOT GUESS! Otherwise, you could really mess things up for your classmates, and they will not appreciate it.
8. If you ask a question and many people respond, summarize all answers and post that summary to benefit your whole class.
9. Be brief. If you write a long dissertation in response to a simple question, it’s unlikely that anyone will spend the time to read through it all.
10. Don’t badmouth others or call them stupid. You may disagree with their ideas, but don’t mock the person.

11. If you refer to something your classmate said earlier in the discussion, quote just a few key lines from their post so that others won't have to go back and figure out which post you're referring to.
12. Before asking a question, check the class FAQs or search the internet to see if the answer is obvious or easy to find.
13. Check the most recent comments before you reply to an older comment since the issue might have already been resolved or opinions may have changed.
14. Be forgiving. If your classmate makes a mistake, don't badger him or her for it. Just let it go – it happens to the best of us.
15. Run a spelling and grammar check before posting anything to the discussion board. It only takes a minute and can make the difference between sounding like a fool and sounding knowledgeable.

Campus Help for Students

Disability-Related Equal Access Accommodations Statement

Students wishing to request academic accommodations to ensure their equitable access and participation in this course should notify the instructor as soon as they are aware of their need for such arrangements. Authorizations from Services for Students with Disabilities (SSD) are generally required. We encourage you to contact SSD at (607) 777-2686 to schedule an appointment with the Director or Learning Disabilities Specialist. The SSD website (www.binghamton.edu/ssd/) includes information regarding their Disability Documentation Guidelines. The office is located in UU – 119.

University Tutoring Services

UTS offers free tutoring for undergraduate students at Binghamton University. All UTS tutoring appointments must be scheduled online through the my.binghamton.edu portal. Walk-in tutoring is also available for select courses. If you have any questions about UTS, call 607-777-9235, email uts@binghamton.edu, or visit the website: http://www.binghamton.edu/tutoring.

ITS Helpdesk/myCourses Support

Walk-in: Located in the Computer Center's first-floor lobby.
Call: 607-777-6420; E-mail: helpdesk@binghamton.edu.
https://www.binghamton.edu/its/

Libraries

The Libraries offer a wide variety and range of services including research assistance, instruction, user-friendly interfaces, digital preservation, digital scanners, and resource sharing.
Text: 607-205-8173; Call: 607-777-2345; Email: refquest@binghamton.edu

http://www.binghamton.edu/libraries

Dean of Students

If you are experiencing undue personal or academic stress at any time during the semester or need to talk with someone about a personal problem or situation, I encourage you to seek support as soon as possible. I am available to talk with you about stresses related to your work in my class. Additionally, I can assist you in reaching out to any one of a wide range of campus resources, including:

1. Dean of Students Office: 607-777-2804
2. Decker Student Health Services Center: 607-777-2221
3. University Police: On-campus emergency, 911
4. University Counseling Center: 607-777-2772
5. Interpersonal Violence Prevention: 607-777-3062
6. Harpur Advising: 607-777-6305
7. Office of International Student & Scholar Services: 607-777-2510

University Counseling Center

At some point during their college experience, students may encounter personal, social, or developmental issues that call for assistance beyond the advice provided by friends and family. That's where the University Counseling Center (UCC) can help. The UCC provides a variety of free and confidential counseling services delivered by professional counselors. All currently enrolled Binghamton University undergraduate students, graduate students and affiliated entities are eligible to receive these services free of charge. Services and programs available through the center include individual and group counseling, consultation, referral, and psychoeducational programs. For more information or to make an appointment, visit https://www.binghamton.edu/counseling.