Syllabus: Applied Semiparametric Regression
STAT 690STA
Spring 2023

Krista J. Gile

Using data to estimate relationships between predictors and responses is an important task in statistics and data science. When datasets are large, modern methods have been developed that allow us to analyze those relationships without making strong assumptions about those relationships- i.e we can let the data determine how $E(y|x)$ relates to x. The methods are generally referred to as "nonparametric regression." This applied graduate course will focus on learning to use nonparametric regression to analyze data. We will read a book, "Semiparametric Regression with R," and implement / interpret the methods in that book. We will address simple and multiple regression data, binary/count data, and correlated data.

1   Objectives

In the context of a dataset and questions about it:
- Learn to specify nonparametric and semiparametric regression models.
- Learn to fit the models with software - mostly in the mgcv package in R.
- Learn to interpret the results of the model.
- Learn to display the results graphically

2   Diversity

2.1   Axioms
- Axiom 1. Mathematical potential is distributed equally among different groups, irrespective of geographic, demographic, and economic boundaries.
- Axiom 2. Everyone can have joyful, meaningful, and empowering mathematical experiences.
- Axiom 3. Mathematics is a powerful, malleable tool that can be shaped and used differently by various communities to serve their needs.
- Axiom 4. Every student deserves to be treated with dignity and respect.

Adapted from Federico Ardila.

2.2   Diversity Statement

Our class includes people from a wide range of personal and professional backgrounds. It has long been a pillar of academic thought that diversity of perspectives makes better science. In this classroom, we celebrate the unique perspective each person brings to the group. We will each have a different path through learning in this class, and different places where we take our learning after this class. While we are together, we all benefit from each other. You will learn more and more deeply the more you learn from each other. In this class, I encourage you to get to know each other, and to support each others' learning, enriched by our diversity of experiences, including persons of varying age, caste, disability, ethnicity, family status, gender, gender identity, geographic region, health status, national origin, political affiliation, race, religion, sexual orientation, social class, and socioeconomic status.

I also recognize that some people in this class face bias, discrimination, and other unjust practices because of parts of their identities. I condemn the parts of our culture that allow and perpetuate these injustices. It is part of my mission to remove and combat these injustices as much as possible in our classroom. I am working to learn more about how injustice shows up in our educational spaces and to improve my teaching practices accordingly. If you ever have concerns or suggestions about this, I invite you, as you feel comfortable, to talk with me about them. If you are not comfortable talking with me, or would like to make a more formal report, you can also report incidents of bias here:

Office of Equity and Inclusion Bias Incident Report https://www.umass.edu/diversity/fight-hate/bias-incident-report

Statistical inference is a powerful tool that can be used to gain insight into many important problems of our time. Much of the history of statistics, along with the history of much western science, has been driven by cis white men. Often, statistics

has been (mis-)used to support bias and oppression. However statistical methods are a tool that can be used to support understanding in all areas of our lives. It is important that people from all backgrounds use statistics to inform insights into all parts of our world. While all trained statisticians might sometimes be able to provide the "right" answer to a particular question asked, only a diversity of statisticians can ask all the "right" questions.

3 How to find me and our course

- Instructor: Krista J. Gile
- Pronouns: She/her
- Times: MWF 10:10-11
- Location: LGRT 171
- Office: LGRT 1435J, and https://umass-amherst.zoom.us/j/98570084401
- Office Hours: Monday 12-1 (hybrid), and Tuesday 2-3 (remote only)
- E-mail: gile@umass.edu
- Website: Moodle
- Homework Submission: Gradescope, code K37GWV
- Backup Zoom Location: https://umass-amherst.zoom.us/j/91018952260
- Prerequisites: Stat 625 or permission of the instructor. We will use R a lot. Highly motivated undergraduates who have taken 525 are welcome too.
- Textbook: Semiparametric Regression with R (Harezlak, Ruppert, and Wand)
- Supplemental (optional) textbook: Semiparametric Regression (Ruppert, Wand, and Carroll)

3.1 Key Dates

Week	Content	Key dates
Feb 6	overview, penalized splines	
Feb 13	penalized splines	add-drop day Monday
Feb 20	penalized splines	no class Monday, **Reading Reflection Due Feb 22**
Feb 27	additive models	
Mar 6	additive models	
Mar 13	Spring Break	no class
Mar 20	generalized additive models	
Mar 27	generalized additive models	**Project Proposal Due Mar 31**
Apr 3	repeated measures/ grouped data	
Apr 10	repeated measures/ grouped data	
Apr 17	bivariate extensions	no class Monday
Apr 24	bivariate extensions	
May 1	other topics	
May 8	other topics/project presentations	
May 15	project presentations	
May 25	Project report	**Project Report Due 11:59pm**

3.2 Computing

- The computing in this course will be conducted in R, freely available software available for multiple platforms at: www.r-project.org. We may have a few bring-your-laptop-to-class days, in which we will work through R exercises in class, and help everyone keep up with R. If you are completely new to R, there are many good online tutorials that can be helpful for getting started.
- All homework will be turned in and graded on gradescope with code above.

4 Requirements and Evaluation

- Homework 50%
- Reading Reflection 5%
- Project 40%
- Class participation 5%

Letter grades are typically as follows:

F	C	C+	B-	B	B+	A-	A
¡ 74	74+	77+	80+	84+	87+	90+	94+

Since this is a graduate class it is very unlikely that a student who does the work will be graded less than a B.

There is a small chance I may adjust a cut-off slightly in the students' favor.

4.1 Homework

Homework will typically be due 30 minutes before the beginning of class on Wednesday. For homework including code, please turn in your homework as a narrative, addressing the question in the homework. Rmarkdown or similar homework approaches that interweave pieces of text, code, and output are good, as long as you use the code, output, and plots to support the words and summary numbers, rather than expecting the reader to read what you've done or pick out the right numerical answers directly from raw output. Please include all code either with each question or at the end.

Homeworks will be uploaded to gradescope in .pdf format. Here are gradescope's suggestions for creating .pdfs. I recommend doing this through `R` using `Rmarkdwon`.

4.2 Project

This is an opportunity to use the methods we develop in this class in practice in a group of 3 or more. Projects will be based on applying the methods in this class (and/or related methods) to a dataset of your choosing. See more information below.

4.3 Reading Reflection

This is an opportunity for you to reflect more on aspects of statistics as a discipline, STEM, and the interface between statistics and the world. For this assignment, you must read a peer-reviewed academic paper in any discipline one one or more of the following topics:

- Talks about the discipline of statistics and its impact on society or how (specific or general) statistical work has advantaged or disadvantaged different social groups.

- Talks about the process of learning statistics or STEM more generally.

- Talks about differences in the process of learning statistics (or other STEM material) for students from different backgrounds or groups.

You may choose from a list of possible papers, or you may find one of your choosing (and check it with me first).

Read the paper, then write a brief reflection to be turned in including:

- Your name
- The reference for the paper you read (with hyperlink if possible, but not essential)
- A 1/2-1 page typed response to the paper, describing briefly the main points of the paper and your personal reflections on the paper and its contents. No more that 1 page.

4.4 Class Participation

Full marks for class participation do not require specific initiative like asking a certain number of questions or volunteering many answers, but do require attending class, fully participating in group work, answering when called on, and paying attention in class. I believe attending class and participating is critical to learning. I do take attendance and I do care if you are there. Nonetheless, I know we live in uncertain times. Please make every effort to make it to class. If you cannot, please contact me (in advance if at all possible) and we will make arrangements so that you can get what you need to get the most out of class (and get full credit for class participation). Missing one or two classes will not impact your participation grade, as long as you are generally present and engaged with the class.

4.5 Missed Work

I expect all students to take responsibility for keeping up with the material. If you need to miss class or an assignment, please communicate with me to make arrangements. Over the course of the semester, no questions asked:

- The lowest homework grade will be dropped
- Each student may request one 'short term late' homework, to be turned in up to 1 week late.
- Each student may request one 'long term late' homework, to be turned in anytime before midnight on reading day, after the end of classes.

In documented extenuating circumstances, please talk to me if other arrangements are necessary.

4.6 Grading Errors

If you believe there was an error in grading your work, please return the work, along with a written description of the problem (you can do this on gradescope) to me within 1 week of the return of that assignment, and I will re-evaluate.

4.7 Excused Absences

If you miss class, it is your responsibility to find out what you missed, both class material and announcements. If a university-approved conflict such as religious observance, athletic event, field trip, performance, health reasons, family illness, or other extenuating circumstances prevents your full participation in the class, please contact me and we will make reasonable arrangements.

Please do not come to class if you are feeling unwell. You can turn in the material remotely, and we can work out the missed work.

4.8   Collaboration and Academic Honesty

Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst.

Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair.

Chegg, Discord, ChatGPT, and other online help resources: Seeking answers directly from any website is a clear violation of the academic honesty policy, while submitting course materials to these sites or similar ones is a violation of the instructor's copyright. Instructors may be monitoring such websites throughout the semester.

The goal of class is to facilitate learning. Most of your learning will come from struggling with the material. Do not be afraid to struggle.

Some online resources including research or teaching materials published online are an important part of the learning and research process. It is OK and encouraged to find scholarly work to support your study and even your homework. (ChatGPT may also become a helpful resource, kind of like a search engine, to help you find related information on a topic of study, although note that it does not come with any guarantee of accuracy, and should always be read with suspicion and primary sources consulted before believing it.). Talking with classmates can also facilitate learning. Students are encouraged to work together and help each other on homework assignments and, of course, projects. However, all submitted individual work must be your own. A good rule of thumb is that another student or online/print source might describe to you their approach to solving the problem, and even show you their work. You should think about their approach, and make sure you understand it. Then look away from their solution, and write up your own solution from your own understanding. You can ask a classmate how they did a tricky part, then adapt an outline like "I used X method then Y method" to create your own solution.

Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. For more information about what constitutes academic dishonesty, please see the Dean of Students's website (p. 9-10).

4.9   Accommodation Statement

The University of Massachusetts Amherst is committed to providing an equal educational opportunity for all students. If you have a documented physical, psychological, or learning disability on file with Disability Services (DS), you may be eligible for reasonable academic accommodations to help you succeed in this course. If you have a documented disability that requires an accommodation, please notify me within the first two weeks of the semester so that we may make appropriate arrangements. For more information, consult the Disability Services website.

4.10   Pronouns

Everyone in this class should be able to choose the pronouns that best describe them. If you would like me to refer to you using pronouns other than what is on spire, please let me know. I appreciate people helping me if I make pronoun errors.

4.11   Other Resources

- UMass Writing Center makes one on one appointments to help with writing. If you are at all concerned about your writing, please consider visiting the writing center for help, especially with your project write-ups. Writing is an important professional skill that is not the focus of most college classes. I strongly recommend you take every opportunity to get input on your writing. For group projects, the group could make a collective appointment, or each interested group member could make a separate appointment about a particular part of the report.

- Graduate Student Senate has resources to support graduate students in various ways.

- UMass Dean of Students has a list of support resources in the area.

- UMass Center for Counseling and Psychological Health helps students find ways to with stress, anxiety, and overall psychological wellbeing.

- UMass Student Success supports students connecting with others and other resources to find their own paths of success and an excellent college experience.

5   Project

The main purpose of this project is to give you experience using methods related to this class in real (often messy or more complex than textbook) data, related to a topic of interest to you. Projects will consist of a careful data analysis. You may

choose your own dataset, or I can help you choose one. Students with topics of interest looking for teammates will have a chance to share their topics with the class to solicit collaboration.

5.1   Deliverables

- Project Proposal: 1-2 pages including the names of all team members and a description of the proposed project.

- In-Class Presentation: A clear, organized presentation, understandable to your classmates. Each team member must take a turn presenting. Practice practice practice! Leave time for questions. Length determined by the number of groups. Please also submit your slides online after the presentation.

- Project Report (one per group). 8-12 pages describing:

- Self-peer evaluation (one per person) due the same time as the project report.

5.2   Suggestions for Presentations

- Begin with introduction/motivation/overview, then the core of your work and findings, followed by results and conclusions

- Practice your talk many times.

- Some questions that may be helpful (not all questions apply to all project types):

    - What is the main question I am trying to answer?
    - How were the data collected/gathered/sampled?
    - Are there any confounding relationships present?
    - Are there any interactions present?
    - Is your model reasonable?
    - What assumptions is it making?
    - What are the limitations of my analysis (assumptions which may not hold, limitations of the data, etc.)?
    - When are the methods I'm using most useful? When might they be misleading?
    - Be sure to cite sources.

5.3   Suggestions for Reports

- The structural suggestions, and guiding questions from the Presentations suggestions apply

- Be sure to cite sources for all data an other references.

- You will probably do a lot more analysis than you present or write about. Keep in the pieces that tell the most interesting story from your analyses. Leave yourselves enough time to edit carefully (and even go to the writing center if you think it will help).