THE UNIVERSITY OF CHICAGO

Masters in Applied Data Science

Course Syllabus

Linear and Nonlinear Models for Business Application
MSCA 31010
Wednesday, 6 – 9 PM
Winter 2024
Location:

Francisco Azeredo
805.637.7688
fazeredo@uchicago.edu

COURSE DESCRIPTION

The primary objective of the course is to provide a practical introduction to statistical methods that extend beyond linear modeling. The statistical techniques taught in this course will enable students to analyze complex datasets and formulate and solve real-world problems. The techniques include generalized linear modeling such as Binomial, Poisson, Gamma, and Gaussian (Normal) regressions, Additional models including multinomial regression, generalized additive models, copulas, mixed regression models, regime switching models, neural networks, fixed and random effects models will also be covered (time permitting).

Throughout the course, students will learn concepts and fundamentals of statistical inference and regression analysis for linear and non-linear methods by studying the theory, developing intuition, and working through simulated and practical examples.

Students will become proficient in interpreting model results and conducting model selection. Students will also learn the statistical programming language used to construct examples and homework exercises. Examples will be constructed primarily using R, but some examples in Python may be used. Lastly, the course will give the students an opportunity to independently learn an advance modeling technique as part of the final group project. Students will have many opportunities to apply new concepts to real data and develop their own statistical routines.

PREREQUISITES

Required: MSCA 31007: Statistical Analysis

COURSE MATERIALS

The following textbooks are recommended:

- Extending the Linear Model with R, by Julian J. Faraway (F)
- An Introduction to Generalized Linear Models by Annette J. Dobson (D)

• Generalized Linear Models by P. McCullagh and J. A. Nelder (MN)

SOFTWARE

R and Python

LEARNING OBJECTIVES

After completing this course, students will not only develop an in-depth understanding of generalized linear models and other advanced non-linear techniques but also develop the ability to:

- apply non-linear models to practical data problems
- interpret statistical results and perform inference
- independently learn a new model from scratch and apply it to a real world data problem

EVALUATION

Grades for the course will be based on participation (10%), homework assignments (40%), and in-class presentation (40%).

Participation

Students should review the relevant class notes prior to class and prepare questions for discussion in class. Students are strongly encouraged to participate in class discussions and activities. Students attending a class on-line must login to the Zoom meeting using your official University of Chicago account. Your attendance will not be counted otherwise. Students' final grade will depend on class participation (20%).

Homework Assignments

Homework assignments are due at the beginning of the class and must be submitted online. Assignments will include conceptual questions, actual data analysis and programming in R and/or Python. Students must individually submit the raw data, a Rmarkdown or Jupyter notebook file with the code and results, technical, detailed and accurate commentary explaining the results as well as a HTML document with all results. Students are encouraged to work in groups of 2-3, but must individually submit their solutions. Homework assignments will constitute 40% of the final grade for the course.

In-Class Presentation

Students will study an advance statistical method as part of a final class project. Students can choose a non-linear model that is **not** a generalized linear model. Students will work as a team (2-3 students), develop an in-depth understanding of the model and apply it to real-world data. Students will submit a final report containing an initial exploratory analysis, a technical discussion of the model estimation and algorithm, estimation results, prediction and testing. Students will present their results in class and receive feedback from classmates and the instructor. The presentations will occur in the last day of class. Each group will be assigned a day and time to present and will be required to attend all presentations on that day.

Each group will have to electronically submit documentation containing the final presentation, all the raw data, code used to organize and analyze the data and produce the results. The presentation will constitute 40% of your final grade for the course.

GRADING SCALE

The overall performance of each student will be assigned a letter grade based on the grading scale below:

A = 93%–100%, A- = 90%–92%, B+ = 87%–89%, B = 83%–86%, B- = 80%–82% ,C+ = 77%–79%, C = 73%–76%

REQUESTING REGRADING

You can request a review of your final grade within two days from the grade submission by emailing the instructor. Your entire performance in the class including homework grades, quiz grades, participation and final presentation will be reassessed and your grade may increase or decrease. Notice that the grade review process follows the MSc. Analytics' grading guidelines.

DISCLAIMER

All class materials are provided with absolutely no warranty. So please use all class materials with caution. It is best practices to review multiple sources for the code/model when developing a solution to a problem and also to tackle a problem using at least two independent approaches. All views expressed by me during class are mine alone and not those of any institution who I am affiliated with.

COPYRIGHTS AND CONTENT

Lecture notes, handouts, exams, quizzes and any other materials provided during the course are intellectual property and are protected by the United States Federal Copyright Law. Students are allowed to use these materials only for their own private learning. Student must request my written permission before sharing any class materials with others. Selling any course material is forbidden. Students' notes taken during classes are considered derivative work and cannot be shared without the instructor's written permission. Any unauthorized videos, pictures and/or audio recording of lectures or discussions is forbidden. Selling and/or distributing class materials or recordings is forbidden. Unauthorized distribution or sale of course work or derivate work constitute copyright infringement and may expose the student to legal actions.

COURSE SCHEDULE

Summer, Autumn, Winter, and Spring quarters are 9 weeks of instruction, with the 10th week for assessment or course rescheduling. Refer to the university's academic calendar at www.uchicago.edu/academics/calendar/ for quarterly start and end dates.

Important Note: Changes may occur to the syllabus at the instructor's discretion. When changes are made, students will be notified via email and in-class announcement.

ATTENDANCE

Your attendance in-person is required and paramount to your success in this class. Students may miss one class with no penalty if the explanation meets my approval. If a student misses a class without my approval, the student's final grade will be reduced by 10%. The second absence, regardless of explanation, will cost the student 5% on the final grade. The third will reduce the grade by an additional 15%, which translates into a maximum possible grade of about a B-/C+. A student who misses four or more classes will fail.

ACCOMMODATIONS

Students who may need long-term accommodation should contact their Dean of Students at mailto:psd-dos@lists.uchicago.edu.

REMOTE ACCESS AND COURSE RECORDINGS

This course is optimized for an in-person experience, but under some circumstances your instructor may choose to provide you access to recordings of the session or the option to join remotely. If an instructor provides you with a recording or a link to join remotely:

- Do not record, share, or disseminate any course sessions, videos, transcripts, audio, or chats.
- Do not share links for the course to those not currently enrolled.
- Any Zoom cloud recordings will be automatically deleted 90 days after the completion of the recording.

LATE WORK

All assignments must be submitted to this course's Canvas site before the next class (by 5:59PM). If you turn in a late assignment, 25% will be deducted from the total score for each day after the deadline. Assignments turned in more than one week late will not receive credit. You should expect to receive your homework grade within 1 week from submission. You can request to have your homework grade reviewed by emailing the grader and copying the instructor. Notice that the grade review process includes reviewing the entire homework submission which may increase or decrease your grade.

REQUESTING REASONABLE ACCOMMODATIONS

The University of Chicago is committed to ensuring equitable access to our academic programs and services. Students with disabilities who have been approved for the use of academic accommodations by Student Disability Services (SDS) and need a reasonable accommodation(s) to participate fully in this course should follow the procedures established by SDS for using accommodations. Timely notifications are required in order to ensure that your accommodations can be implemented. Please meet with me to discuss your access needs in this class after you have completed the SDS procedures for requesting accommodations.

Phone: (773) 702-6000
Email: disabilities@uchicago.edu

Please follow accommodation implementation instructions provided by the disability liaison in the division after you have completed the SDS procedures for requesting accommodations.

You may want to begin by reading through the information published on this website https://disabilities.uchicago.edu/. Contact SDS at disabilities@uchicago.edu or 773-702-6000, if you are interested in requesting disability accommodations.

ACADEMIC HONESTY & PLAGIARISM

It is contrary to justice, academic integrity, and to the spirit of intellectual inquiry to submit another's statements or ideas of work as one's own. To do so is plagiarism or cheating, offenses punishable under the University's disciplinary system. Because these offenses undercut the distinctive moral and intellectual character of the University, we take them very seriously.

Proper acknowledgment of another's ideas, whether by direct quotation or paraphrase, is expected. In particular, if any written or electronic source is consulted and material is used from that source, directly or indirectly, the source should be identified by author, title, and page number, or by website and date accessed. Any doubts about what constitutes "use" should be addressed to the instructor.

Academic Honest and Plagiarism:
https://studentmanual.uchicago.edu/academic-policies/academic-honesty-plagiarism/
https://internationalaffairs.uchicago.edu/page/honest-work-and-academic-integrity-plagiarism

Copyright: http://www.lib.uchicago.edu/copyrightinfo/

STUDENT HEALTH PACT

All students on campus are required to adhere to the guidelines in the UChicago Health Pact in order to create a safe environment in the classroom. This entails:

- Wearing a face covering or mask covering your nose and mouth in University facilities when required by the University, unless you are not required to wear a face covering or mask because you have received an accommodation;

- Keeping a face covering with you at all times in the event it is needed regardless of COVID-19 vaccination status;
- Washing your hands with soap and water frequently for at least 20 seconds, or using alcohol-based hand sanitizer when soap and water is not available, throughout the day and before and after in-person interactions with others.

The complete text of the UChicago Health Pact along with additional information about COVID-19 protocols can be found here.

REPORTING COVID-19 RELATED CONCERNS

If you believe that a required COVID-19 safety policy or practice is not being followed, you should report the incident to your supervisor, academic leader, or via the University of Chicago Accident Incident Reporting (UCAIR).

REPORTING COVID-19 EXPOSURE OR A CONFIRMED CASE

Individuals who are tested positive for COVID-19, have COVID-19 related symptoms but have not tested positive or have tested negative, been exposed to a COVID-19 positive person and are currently asymptomatic, or recently traveled are to follow the University's Protocol for Addressing Confirmed or Suspected COVID-19 Exposures.

According to the protocol, any University community member who has a confirmed case, symptoms, or exposure must promptly self-report via C19HealthReport@uchicago.edu and cooperate with the Contact Tracing Team.

RECORDING AND DELETION POLICIES FOR ACADEMIC YEAR 2020-2021

The Recording and Deletion Policies for the current academic year can be found in the Student Manual under Petitions, Audio & Video Recording on Campus.

- Do not record, share, or disseminate any course sessions, videos, transcripts, audio, or chats.
- Do not share links for the course to those not currently enrolled.
- Any Zoom cloud recordings will be automatically deleted 90 days after the completion of the recording

LECTURES

LECTURE 01
(F: Ch. 1; D: Ch. 1)

Provides introduction, scope and preview of the course. Revisits the linear model framework and discuss its extensions. Motivates the use of non-linear models in business analytics. Explores different data types, responses and explanatory variables and statistical methods. Introduces generalized linear models and their key components (response, regressors, link function, linear predictor, response distribution), estimation method (maximum likelihood) and goodness of fit methods (deviance, AIC). Reviews the standard linear regression model in the context of generalized linear models.

LECTURE 02
(D: Ch. 3)

Introduces the exponential family of distributions and the generalized linear model framework as defined by Nelder and Wedderburn (1972). Describes properties of the exponential family of distributions. Discusses the generalized linear model and its key components, including: density, join density, link function, parameterization, and response and explanatory variables. Presents standard examples including Poisson, Binomial, and Multinomial models.

LECTURE 03
(D: Ch. 4)

Outlines and discusses maximum likelihood estimation in the context of univariate distributions and generalized linear models. Presents numerical methods for finding point estimates for generalized linear model parameters, including Newton-Raphson algorithm. Numerical examples are presented and discussed in detail.

LECTURE 04
(D: Ch. 5)

Presents methods for statistical inference for generalized linear models. Presents statistical inference methods, including confidence interval estimation, hypothesis testing and goodness of fit measures (e.g., deviance, AIC, BIC). Considers the use of asymptotic results under non-normality assumptions.

LECTURE 05
(D: Ch. 6)

Discusses the normal linear model. Examines maximum likelihood estimation, link function, statistical diagnostics, goodness of fit and prediction. Discusses simulated and practical examples and conducts in-class group exercises.

LECTURE 06
(F: 2; D: Ch. 7)

Explores the binomial model. Examines maximum likelihood estimation, logistic regressions, link functions, statistical diagnostics, goodness of fit measures, and predictive analytics. Comparison of different binomial regression model with logit, probit and complementary log-log link functions. Discusses simulated and practical examples and conducts in-class group exercises.

LECTURE 07
(F: 3; D: Ch. 9)

Explores the Poisson regression and negative binomial models. Examines maximum likelihood estimation, Poisson and negative binomial regressions, link functions, statistical diagnostics, goodness of fit measures, and predictive analytics. Discusses simulated and practical examples and conducts in-class group exercises.

LECTURE 08
(F: 5; D: Ch. 8)

Explores the multinomial regression model. Examines maximum likelihood estimation, nominal and ordinal logistic regressions, link functions, statistical diagnostics, goodness of fit measures, and predictive analytics. Reviews the negative binomial distribution as a Gamma-Poisson mixture distribution. Discusses simulated and practical examples and conducts in-class group exercises.

LECTURE 09
(F: 7; Notes)

Explores the gamma regression model. Examines maximum likelihood estimation, nominal and ordinal logistic regressions, link functions, statistical diagnostics, goodness of fit measures, and predictive analytics. Discusses simulated and practical examples and conducts in-class group exercises.

LECTURE 10
(Notes)

Explores some special topics in non-linear modeling (time permitting), including: copulas, mixed regression models, regime switching models, neural networks, generalized additive models, models with fixed, random and mixed effects.