



School of Professional Studies

Course Name and Number: DATA 621 - Business Analytics and Data Mining

Spring 2017

Credits: 3 cr.

Prerequisites: DATA 606 - Statistics and Probability for Data Analytics; DATA 607 - Data Acquisition and Management Course

My Contact Information:

Instructor Name: Sabrina Khan

E-mail Address: Sabrina.khan@sps.cuny.edu

Office Hours / Sync: Online only. By appointment.

Phone: (929)-245-0383

Description: This course develops the foundations of predictive modeling by introducing the key concepts of applied regression modeling and its extensions. The main topics covered in this course include: simple and multiple linear regression, variable selection and shrinkage methods, binary logistic regression, count regression, weighted least squares, robust regression, generalized least squares, multinomial logistic regression, generalized linear models, panel regression, and nonparametric regression. The course is heavily weighted towards practical application using the R statistical programming language and data sets containing missing values and outliers. The course also addresses issues of exploratory data analysis, data preparation, model development, model validation, and model deployment.

Course Learning Objectives:

By the end of the course, students should be able to:

- Demonstrate a practical understanding of the theoretical concepts behind applied regression modeling.
- Analyze and select appropriate types and combinations of models given particular business situations.
- Develop applied regression modeling techniques to address different types of data.
- Use R statistical software to build and deploy specific models based on real-world business problems.

Program Learning Outcomes/Competencies addressed by the course:

- Business Understanding. Students will learn how applied regression modeling techniques can add value to existing business analytics.
- Data Programming. Use industry standard statistical programming tools.
- Foundational Math and Statistics. Emphasis on probability, statistics, and computational methods.

- **Data Culture.** Students will learn how applied regression modeling can enhance business capabilities and extend the value of existing data.
- **Data Understanding.** Students will learn how to explore data to find new patterns.
- **Predictive Modeling.** Selecting predictive modeling techniques, building and assessing models.
- **Model Implementation.** Students will learn to implement models for the various applied regression modeling techniques covered in the course.

How is this course relevant for IS and data analytics professionals?

Regression modeling skills are crucial, high-value skills in today's data-driven business environment where real-world decision-making processes are complex. The ability to leverage rapidly expanding data sets to obtain new insights is at the heart of predictive data analytics.

How does this course work?

The course is conducted entirely online via Blackboard. Each week, the student will complete assigned readings from the required textbooks, watch lecture videos, complete optional (but recommended) textbook exercises, complete homework assignments (not weekly), and participate in the discussion board. There is also a final course project. Students are expected to complete all deliverables by their assigned due dates.

Assignments and Grading:

Homework Assignments - There will be 5 homework assignments (10% each, or 100 points each) used to re-enforce course concepts and provide implementation experience in R. - Students may collaborate with their classmates on the homework but the final submission needs to be one's own work.	50%	500
Final Project - Students will form a group of 3-4 people. - Each group will submit a project. - The project will require the students to model a problem using any of the methods learned in this course.	30%	250
Class discussion - Each week, we will have a topic to discuss. A student is required to participate either by introducing a question or by answering someone else's question. - A total of 5 blog entries will be required throughout the semester based on a topic of your choice where a student shares his or her thoughts on a statistical method and how that can be used in a real life scenario based on your daily life experience. These need to be completed by the end of the semester.	20%	200
TOTAL	100%	1000

Grading Scale: Your grade will be based on your final weighted average score and the letter grade will be assigned according to the following table. However, depending on how the class does, we will see if a different scoring schema may be a more appropriate approach.

Letter Grade	Range %	GPA
A	93-100	4
A-	90-92.9	3.7
B+	87 - 89.9	3.3
B	83-86.9	3.0
B-	80-82.9	2.7
C+	77-79.9	2.3
C	70-76.9	2.0
F	<70	0.0

Discussion Board Etiquette: The purpose of the discussion board in general is to allow students to freely exchange ideas. It is imperative to remain respectful of all viewpoints and positions and, when necessary, agree to respectfully disagree. While active and frequent participation is encouraged, cluttering a discussion board with inappropriate, irrelevant, or insignificant material will not earn additional points and may result in receiving less than full credit. Frequency is not unimportant, but content of the message is paramount. *Please remember to cite all sources (when relevant) in order to avoid plagiarism.*

Late Policy: Unless otherwise noted, all work is due on the assigned day by 11:59 PM (Eastern Time). This includes homework assignments, projects, and participation in the discussions. *In case of an extenuating circumstances, we can make exception. Please be sure to contact me ahead of time.*

Required Textbooks:

- *A Modern Approach to Regression with R*, by Simon J. Sheather. ISBN 978-0-387-09608-7 **(MARR)**
- *Linear Models with R*, by Julian J. Faraway. ISBN 978-1439887332 **(LMR)**
- *Extending the Linear Model with R*, Julian J. Faraway. ISBN 978-1584884248 **(ELMR)**

Relevant Software: The primary software environment is the R statistical programming language, which can be downloaded for *free* from <http://www.r-project.org>. RStudio is the recommended interface for the R statistical programming language software, which can also be downloaded for *free* at <http://www.rstudio.org>.

Note on timing of communications:

Due to working in a restricted environment during the day, I will dedicate sometime every weeknights for responding to emails. Weekends will be typically reserved for conducting all time-consuming work (e.g. grading, responding in more details, if anyone needs to schedule some time to talk/meet via Skype, etc.). Therefore, most assignments will be due on Saturday. You are encouraged to ask questions on the course discussion board where other students will be able to benefit from your inquiries. For the most part, you can expect me to respond to questions by email within 24 to 48 hours. If you do not hear back from me

within 48 hours of sending an email, please resend your email. You can expect me to grade and return assignments within 14 days. Please do not hesitate to ask if you have questions or concerns.

Tentative Course Outline:

Please note that this schedule is subject to change depending on our progress, questions, requests, etc.

Week	Topic	Key Task
Week # 1 Jan 30 – Feb 5	Introduction to Applied Regression Modeling	Discussion # 1 due
Week # 2 Feb 6 – Feb 12	Exploratory Data Analysis	Discussion # 2 due
Week # 3 Feb 13 – Feb 19	Simple Linear Regression: Estimation	Discussion # 3 due Homework # 1 assigned
Week # 4 Feb 20 – Feb 26	Simple Linear Regression: Inference, Prediction, Explanation	Discussion # 4 due
Week # 5 Feb 27 – Mar 5	Simple Linear Regression: Diagnostics and Transformation	Discussion # 5 due Homework #1 due Homework #2 assigned
Week # 6 Mar 6 – Mar 12	Multiple Linear Regression and Missing Data	Discussion # 6 due
Week # 7 Mar 13 – Mar 19	Multiple Linear Regression: Variable Selection	Discussion # 7 due
Week # 8 Mar 20 – Mar 26	Multiple Linear Regression: Model Diagnostics	Discussion # 8 due Homework # 2 due Homework #3 assigned
Week # 9 Mar 27 – Apr 2	Logistic Regression	Discussion # 9 due
Week # 10 Apr 3 – Apr 9	Logistic Regression for Rare Event	Discussion # 10 due Homework #3 due Homework #4 assigned
Week # 11 Apr 10 – Apr 16	Spring Break	
Week # 12 Apr 17 – Apr 23	Generalized Least Squares	Discussion # 11 due Homework #4 due Homework #5 assigned Project assigned Form project teams
Week # 13 Apr 24 – Apr 30	Multinomial Logistic Regression	Discussion # 12 due
Week # 14 May 1 – May 7	Generalized Linear Models	Discussion # 13 due
Week # 15 May 8 – May 14	Repeated Measure and Longitudinal Data	Discussion # 14 due Homework # 5 due
Week # 16 May 15 – May 21	Nonparametric regression	Discussion # 15 due
Week # 17 May 22 – May 26		Project report due All blog entries due

ACCESSIBILITY AND ACCOMMODATIONS

The CUNY School of Professional Studies is firmly committed to making higher education accessible to students with disabilities by removing architectural barriers and providing programs and support services necessary for them to benefit from the instruction and resources of the University. Early planning is essential for many of the resources and accommodations provided. Please see: http://sps.cuny.edu/student_services/disabilityservices.html

ONLINE ETIQUETTE AND ANTI-HARASSMENT POLICY

The University strictly prohibits the use of University online resources or facilities, including Blackboard, for the purpose of harassment of any individual or for the posting of any material that is scandalous, libelous, offensive or otherwise against the University's policies. Please see: http://media.sps.cuny.edu/filestore/8/4/9_d018dae29d76f89/849_3c7d075b32c268e.pdf

ACADEMIC INTEGRITY Academic dishonesty is unacceptable and will not be tolerated. Cheating, forgery, plagiarism and collusion in dishonest acts undermine the educational mission of the City University of New York and the students' personal and intellectual growth. Please see: http://media.sps.cuny.edu/filestore/8/3/9_dea303d5822ab91/839_1753cee9c9d90e9.pdf

STUDENT SUPPORT SERVICES If you need any additional help, please visit Student Support Services: http://sps.cuny.edu/student_resources/