2020

MATH 310: Applied Regression Analysis

Yu Wang
CUNY York College

MATH 310: Applied Regression Analysis (Hybrid/Online)

Fall 2020 Section: NO

Instructor: Dr. Yu Wang

Location: AC/CLB or Blackboard Collaborate

Time: T/Th 10:00 – 11:50 am

Office: AC/2D03

Phone: 718-262-2547

Email: vwang AT York.cuny.edu

Office Hours: T/TH 9:00-9:50 (AC/2D03 / Blackboard Online) or by appointment

Teaching Philosophy/Portfolio: https://commons.gc.cuny.edu/members/vwang/

Prerequisites: ***Math 210*** Probability and Inferential Statistics I.

Course Description

Introduce the different linear statistical models and develop critical thinking for statistical modeling in scientific and policy contexts; Apply statistical computer software tools to develop useful data analysis skills based on the use of linear regression models. Topics to be covered: simple linear regression, multiple regression, nonlinear regression and logistic regression models; Random and mixed effects models; The application of statistical software tools.

Credits: This is a 4-credit class.

Course Narrative

This is an **Open Education Resource** (**OER**)/ **Zero Textbook Cost** (**ZTC**) course. The textbook and software package **R** can be downloaded online free. The lecture notes, homework assignments, and project reports will be added to OER websites for similar courses use. Moreover, we will apply **Active Learning and OER** Pedagogy in this class, as the following:

1. The active learning pedagogy application to class: Partial class notes will be provided EXCEPT examples and Q & A on the free online platform MyMathOpen.com / CUNY Commons. You can focus on working examples instead of copying notes of definitions and methods. Some examples can be completed before class independently or with discussion, while some advanced examples will be discussed in class. Each student will have opportunity to present your OWN solution of examples and problems in class. As a

result, class notes with all examples completed in class will be **shared** and **editable** on MyMathOpen / CUNY Commons for future students who will have more open examples to learn with any OER course.

2. All the problems of assignments are from OER, which will be posted and graded through MyMathOpen / CUNY Commons. There are two categories of the assignments, one is individual and the other is group part. You need finish the problems in the individual part first based on the examples we have discussed in class and then work on a challenging one in the group part, which can be **discussed** and **shared** using active learning pedagogy. Finally, you can **CREATE** one or two bonus worked examples for sharing and downloading on MyMathOpen.com, so that we can contribute materials to the Open Education Resources.
3. When the data scale is large, two statistical software package R-focused projects will be given during the semester, which also require the active learning and open pedagogy. The first project is a single variable regression model. You need to apply the knowledge discussed in class and write some simple R code to run regression analysis. Finally, you will summarize the result from R and submit your conclusion individually. (Two weeks) The second project with active learning is a teamwork project. Each group will apply active learning pedagogy to collect data, plot related graphs, analyze the fitness of the regression model you select, and organize all in a written report step by step (Three to Four weeks) with skills you obtain from Project 1 and all homework assignments. In addition, the conclusion is open and not unique, which assist you better understand the core idea of active learning and the cooperation of teamwork, and the procedure of various tasks which will help you understand and implement any related data analysis project in the future. The method and procedure of completing the project will benefit all other OER learners as well. The final report (without your original R code) will be posted and editable on MyMathOpen / CUNY Commons for future OER courses to use and generalize.

Textbook: Linear Regression Using R: An Introduction to Data Modeling (PDF) by David J. Lilja, University of Minnesota, ISBN: 13: 9781946135001

https://open.umn.edu/opentextbooks/textbooks/linear-regression-using-r-an-introduction-to-data-modeling

Learning Objectives:

1. Review the concepts of probability distributions and confidence intervals;

2. Understand the basic linear regression models in statistics;

3. Identify different regression models and apply different methods to build and fit the models;

4. Discuss the validation and subsequent inferences for various models;

5. Acquire skills to identify real problems that require advanced mathematical and statistical knowledge;

6. Learn and apply statistical software package R to linear regression models for data analysis.

Ancillary Learning Objectives:

- Acquire skills to write report of statistical inference with detailed programming code;
- Understand and complement the whole procedure of collecting, processing, analyzing, and summarizing data;
- Learn the collaboration skills by completing some computer projects of team work.

Class Supplies: Students are required to bring a notebook, pens/pencils and a scientific calculator or computer (not necessarily graphic calculator) to class (Online/f2f) each day. You may not use a calculator with memory or the calculator on your Cell Phone/PDA for tests.

Course Requirements

Homework Assignments: Homework will be assigned on MyOpenMath/Blackboard each class. It is to be turned in, graded and returned to you (Online). Late homework is not accepted and a grade of zero is assigned for missing assignments. Each homework assignment will consist of problem sets: statistical problems related to the work discussed in class. Problems will be reviewed in class as needed. Problem sets will be graded on a scale of 0 - 100 points.

Examinations / Makeups: There will be ONE in-class/online Midterm exam and One in-class/online Final Exam on topics covered in class throughout the semester. Exams will consist of statistics problems of the type discussed in class sessions and through homework assignments. In general, there will be NO MAKEUP exams including final exam.

Projects: The are two computer-based projects using **R** in total. The first project is an independent project. You are required to submit the summary of the analysis (up to 10 pages), the original programming code, and the necessary plots. The second project is a group project. Up to 4 students can complete the project in one group and submit a final report of the project with the original programming code and all necessary outputs and plots from the computer software package.

Sample Project

The following is a sample of Project 2 - a group work, which will be an application of combined knowledge we learn in the whole semester. Meanwhile, it is a testification of Active Learning Pedagogy of our class and it will count 20% of the overall grade. After you complete

the project, the report without the original R code will be posted on MyMathOpen.com / CUNY Commons as an OPEN EDUCATION RESOURCE material which is one contribution to OER:

Assessing the worth of a diamond stone is no easy task in view of the four C's, namely caratage, color, clarity and cut. (Our) objective is to infer a sensible pricing model for diamond stones based on data pertaining to their weight (in carats), their color (either D, E, F, G, H or I) and clarity (either IF, VVS1, VVS2, VS1 or VS2). Of interest is the relative worth of the different grades of color and clarity and whether differences in prices can be attributed to the 3 different certification bodies (either GIA, IGI or HRD).

Columns

1 Carat - Weight of diamond stones in carat units
2 Color - D, E, F, G, H or I
3 Clarity - IF, VVS1, VVS2, VS1 or VS2
4 Certification Body - GIA, IGI or HRD
5 Price (Singapore $)

SOURCE: The data appeared in Singapore's _Business Times_ edition of February 18, 2000.

Suppose you are asked for the best model to predict Price based on the above dataset using R. Come up with the "best" model to show the person. You may not decide on the same model as other students!

Each time you fit the model, follow these steps. Outline them clearly in your report just like you did last time. Make sure you comment on every step!

In this group project, 2-4 members will follow these active learning steps:

1. Discuss the goal of the project and allocate different tasks to each member, like collecting and organizing data, writing R code, analyzing result of R, and summarizing the result;
2. Collect, classify, and organize the data set;
3. Review all the multi regression models in this course and select all possible models which can be used in this project;
4. Write R code to test different regression models step by step;
5. Analyze the results of regression models in R using the knowledge you gain from Project 1 and all available problems in class;
6. Select the "best" model based on the results you think in Step 5 and interpret your selection;
7. Summarize the whole procedure and write each step in a final report.

You can use all class notes, OER textbooks, and all methods we discuss in class and completed problems. Please notice that all the above procedures of this project will be provided as an OER material for future students to be generalized and completed if necessary.

All submitted examples, problems, and project reports will be shared with a Creative Commons license. A CC license allows users to copy, share, and build upon the work. In return, you will also have the chance to similarly copy, share and remix other submissions that are shared with a CC license. Please take a look at Creative Commons for more information about the types of license that we will consider as a class for this project.

Grading Breakdown

The following will be used to determine a student's grade in the course. Percentagewise, the breakdown of the grade is as follows:

Homework Assignments 20%

First Project Score 15%

Second Project Score 20%

Midterm Exam Score 20%

Final Exam Score 25%

Course and Instructor Policies

Attendance will be recorded each class (Online/f2f) for final record. Class participation is very important. Each forum post will earn bonus points (up to 10 in total). Communication related to course in class (Online/f2f) is necessary and encouraged. All emails and online questions will be responded in 24 hours (weekdays) to 48 hours (weekends/holidays).

Special Dates

Monday, September 7	College is closed, no classes scheduled
Fri.- Sun. September 18-20	No classes scheduled
Monday, September 28	No classes scheduled
Tuesday. September 29	Classes follow a Monday schedule
Monday, October 12	College is closed, no classes scheduled
Wednesday, October 14	Classes follow a Monday schedule
Wednesday, November 25	Classes follow a Friday schedule
Thurs- Sun, November 26-29	College is closed, no classes schedule

Important Policies

Policy on Academic Integrity York College's Academic Integrity Policy and Procedures, developed to conform to the CUNY policy on Academic Integrity, outline College protocol for (1) promoting academic integrity at the College; and (2) dealing with violations of academic integrity. Academic Dishonesty is prohibited in The City University of New York and is punishable by penalties, including failing grades, suspension, and expulsion. Academic dishonesty includes cheating and plagiarism. This policy will be strictly adhered to. Students can familiarize themselves with this policy by downloading a copy of it in pdf form at http://york.cuny.edu/president/legal-compliance/legalaffairs/cuny-legal-policies-procedures/Academic-Integrity-Policy.pdf/view

Policy on INC grades York Colleges policy on INC grades will be adhered to. Students should familiarize themselves with the policy. Additionally York Colleges grading scale will be used to determine final course grades. All of Yorks grading policies can be found at: http://york.cuny.edu/academics/policies/grading-policies

Policy on accommodations for disabled students York Colleges policy on accommodations for disabled students will be strictly adhered to. Students with documented disabilities are entitled to receive accommodation, including in some cases extra time on exams, tests, projects and assignments. The Office of Services for Students with Disabilities located within the Counseling Center in AC/1G02 provides a wealth of support and services including accommodations such as: extended testing time, large print text, adjustable tables, and computers services provided for by OSD. To better ascertain if you are eligible for services please stop by and speak with a disability's specialist. The Office of Student Disabilities (Room AC-1G03) can evaluate students. More information is also available online at http://york.cuny.edu/student-development/ossd/

Policy on Student Evaluations Department policy requires that the following statement must be included in all department syllabi. All student evaluations of teaching must be given online during the semester. Check the email for the access information.

Student Support Resources

- IT Support: 718-262-5300
- Blackboard and Technologies support: Check Blackboard for more details
- Library Reference Desk: AC Floor 3 at the entrance of the library
- OER/ZTC Support: Katherine Tsan: ktsan@york.cuny.edu

Note Any change to this syllabus will be announced in class.

Topics and Approximate Schedule

The following table outlines the topics that will be covered as part of the course and a timeline for their completion. Please note that the above schedule is approximate and is subject to change.

Date	Topics of Study	Reading	Notes
Week 1 (f2f)	Course motivation; Introduction to linear regression analysis; Review of statistical inference: sampling distributions, tests and confidence intervals; computer packages	1.1 - 1.2	
Week 2 (Online)	Two-sample t-tests, one-way ANOVA, multiple comparison procedures	1.3 - 1.5	
Week 3 (Online)	Simple linear regression, tests and confidence intervals for slope and intercept, prediction, model assessment	2.1 - 2.6	HW of Chapter 1
Week 4 (f2f)	Correlation models; R2 and the ANOVA table	2.7 - 2.10	
Week 5 (Online)	Multiple regression and inferential tools for multiple regression	3.1 - 3.11	HW of Chapter 2
Week 6 (Online)	Regression Diagnosis	4.1 - 4.14	
Week 7 (Online)	Dummy variables, two-way ANOVA, ANCOVA	5.1 - 5.7	HW of Chapter 3
Week 8 (f2f)	Review, Midterm Exam in class		Project 1
Week 9 (Online)	Regression with transformed variables, polynomial regression	6.1- 6.9	HW of Chapter 4, 5
Week 10 (Online)	Weighted least squares, serial correlation	7.1 - 7.5	
Week 11 (f2f)	Multicollinearity; bias variance tradeoff, penalized regression	8.1 - 8.10	HW of Chapter 6
Week 12 (Online)	Variable selection; non-linear regression	9.1 - 9.9	Project 2
Week 13 (f2f)	Generalized linear models	10.1 -10.8	HW of Chapter 7
Week 14 (f2f)	Review for Final Exam		HW of Chapter 8, 9
Finals Week (f2f)	Final Exam as per York schedule		Final Project Due