HUDM 5122-Applied Regression Analysis
Fall 2024 Syllabus
Teachers College, Columbia University

Instructor: Youmi Suk
Email: ysuk@tc.columbia.edu
Office: 552 Grace Dodge
Office Hours: Tue. 4:00-5:00pm, In-Person. Online sessions available upon request.

Teaching Assistant: Weixuan Xiao
Email: wx2299@tc.columbia.edu
Office Hours: Fri. 2:30-3:30 pm,
Online

Lecture: Tue. (In-Person) & Thu. (Online) 5:10-6:50pm, GD 281 or Zoom
Note that an online lecture will be offered on Thu, Sep 5-th.

Course Website: Canvas

Course Overview and Learning Outcomes:
HUDM 5122 is an introduction to regression with emphasis on the practical aspects. Topics include: simple linear regression, multiple linear regression, regression with categorical independent variables and/or interactions, role of assumptions, model diagnostics, and generalized linear models.

At the end of the course, students will

(1) Be able to correctly choose and apply common regression methods that are used in practice to analyze data, including simple and multiple linear regressions as well as generalized linear models (e.g., logistic).
(2) Correctly interpret and explain results from regression methods, including interpretation of the coefficients, the p-values, R-squared, and other statistical summaries from regression.
(3) Understand the underlying assumptions behind common regression methods and utilize diagnostic tools to detect violations of said assumptions.
(4) Apply these methods to real data using the statistical software SPSS.

Prerequisites:
The official prerequisite for the course is HUDM 4120, HUDM 4122 or equivalent.

Textbook:
The primary textbooks are Field (2009) and Allison (1998). All of the course materials will be based on combinations of the below textbooks.

- Field, A. (2009) Discovering statistics using SPSS. Third edition. SAGE.
- Allison, P. (1998) Multiple regression: A primer. The Pine Forge Press series in research methods and statistics. Thousand Oaks, CA: Sage. (available on Columbia Library and at the Columbia Bookstore)
- Hardy, M. (1993) Regression with dummy variables. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-093. Newbury Park, CA: Sage. (available on Columbia Library EResources)
- Fox, J. (1991) Regression diagnostics. Sage University Papers Series on Quantitative Applications in the Social Sciences, 07-079. Newbury Park, CA: Sage. (available on Columbia Library EResources)

Software:
IBM SPSS Statistics. SPSS is free to install and use for TC students and faculty through the myTC Portal. To access SPSS,

- Go to Student/Faculty Resources on the myTC portal.
- Find Academic Technology Resources.
- Click SPSS Home Use for Students and Faculty.

Data
For your final project, you will analyze real data and draw meaningful conclusions with regard to your research questions. These days it is actually fairly easy to get your hands on interesting data. Here is a list of websites where you can find interesting data.

- kaggle
- AWS Open Data
- data.world
- ICPSR
- The Google Dataset Search
- The UCIML Repo
- The CMU data repository
- The datasets subreddit
- Tycho
- Data Portals

Class Management
Email / Communication
- For course-related questions, email is the best way to contact the teaching staff, including the professor and TA. Please be sure to include "HUDM 5122" in the subject line of your email.
- If you have concerns about the course schedule or assignments, please inform the instructor within the first three weeks of the semester.
- The teaching staff will not respond to inquiries unrelated to this course, such as requests for journal paper revisions, assignments from other courses, personal readings, or similar tasks.

Course Schedule

The following calendar is an outline of the course materials for HUDM 5122.

Week/Date	Topics	Reading (before lecture)	Homework (due Monday)
Week 1 Sep 3 & 5	Syllabus, Introduction, Descriptive statistics review (Lecture and Lecture)	Field Ch 1 & 4	Pre-class survey (due Sep 9; ungraded)
Week 2 Sep 10 & 12	Statistics review (Lecture and Lab)	Field Ch 2 & 3	HW #1 {Review} (due Sep 16)
Week 3 Sep 17 & 19	Simple linear regression model (Lecture and Lab)	Field Ch 7.1-7.4 Allison Ch 1 & 5	HW #2 {Simple reg.} (due Sep 23)
Week 4 Sep 24 & 26	Multiple linear regression model (Lecture and Lab)	Field Ch 7.5-7.10 (Skip 7.6.1-7.6.2 & 7.8.5-7.9) Allison Ch 2 & 3	HW #3 {Multi. reg.} (due Sep 30)
Week 5 Oct 1 & 3	Regression with categorical independent variables (Lecture and Lab)	Field Ch 7.11 Hardy Ch 1–3 (pp. 1-28)	HW #4 {Dummy var} (due Oct 7)
Week 6 Oct 8 & 10	Regression with interactions (Lecture and Lab)	Allison Ch 8 Hardy Ch 4 (pp. 29-63)	HW #5 {Interactions} (due Oct 14)
Week 7 Oct 15 & 17	Regression with non-linearity (Lecture and Lab)	Allison Ch 8 Hardy Ch 6.2-6.3 (pp. 78-82)	HW #6 {Non-linearity} (due Oct 28)
Week 8 Oct 22 & 24	Midterm (Oct 22, in-person), Midterm review (Lab)		Midterm
Week 9 Oct 29 & 31	Real data analysis: datasets, codebooks, cleaning/importing (Lecture and Lab)		HW #7 {Real data} (due Nov 4) Groups (due Oct 31)
Week 10 Nov 7	College Holiday (Election Day) Group meetings for the final project (Lab only)		Project Proposal (due Nov 11)
Week 11 Nov 12 & 14	Model diagnostics 1: Outliers, leverage, and influence (Lecture and Lab)	Fields Ch 7.6.1, 7.8.6 Allison Ch 6 Fox Ch 4	HW #8 {Outliers, etc} (due Nov 18)
Week 12 Nov 19 & 21	Model diagnostics 2: Non-normality, Non-constant variance, Collinearity (Lecture and Lab)	Field Ch 7.6.2, 7.8.5, 7.8.7, & 7.9 Allison Ch 7 Fox Ch 3, 5, 6, & 10	HW #9 {Non-norm.} (due Nov 25)
Week 13 Nov 26	Generalized linear model (logit model) (Lecture and recorded Lab)	Field Ch 8 Hardy Ch 6.1 (pp. 75 – 78)	HW #10 {Logistic 1} (due Dec 2)
Week 14 Dec 3 & 5	Generalized linear model (logit & probit model) (Lecture and Lab)		Presentation slides (due Dec 9)
Week 15 Dec 10 & 12	Project presentation (Dec 10) Quiz (Dec 12, in-person)		
Week 16	No class; Final project due		Final Project (due Dec 17)

Note:
Hardy Ch 1–3 (pp. 1-28): Introduction; Creating Dummy Variable; Using Dummy Variables As Regressors
Hardy Ch 4 (pp. 29-63): Assessing Group Differences in Effects
Hardy Ch 6.2-6.3 (pp. 78-82): *Testing for Curvilinearity*; *Piecewise Linear Reg*.
Hardy Ch 6.1 (pp. 75-78): *Dummy Variables in Logit Models*.
Fox Ch 4: Outlying and Influential Data
Fox Ch 3, 5, 6, & 10: Collinearity; Non-Normally Distributed Errors; Nonconstant Error Variance; Recommendations

Weekly Assignments, the Midterm, the Quiz, and the Final Project

Weekly Assignments: The weekly assignments consist of focused exercises related to each week's lectures (e.g., interpreting regression coefficients, assessing assumptions). The assignments will be uploaded on Canvas each Tuesday and will be due the following Monday at 11:59pm, EST; see each due on the course schedule. Our TA will view students' submissions and make comments on them. Assignments are expected to be completed by due date. The total possible points for each assignment will vary. Assignments that are turned in late will be subject to the following penalty: 10% of the total score will be deducted for every day. An assignment with the lowest score will be dropped when computing the final score at the end of the semester.

Midterm: The midterm will cover Weeks 1–6, and will be an in-person exam. You will be allowed one 8 ½ x 11 one-sided "cheat sheet." You're allowed to bring a calculator.

Quiz: The quiz will cover Weeks 11-13, and will be an in-person, open-book test. You're allowed to bring a calculator.

Final Project: In the "real world" of academia, empirical papers are rarely written by one person. Usually, papers include 2-4 people. Thus, you will work in groups of 2-4 students. A great team will include people with different assets. Some may be better at statistics, while others are better at writing, while others know the subject material better. You must let TA know who exactly is in your group by *Oct 31* (Thu). Do this sooner rather than later, so that you can get started working on your dataset. Each student in the group must (a) communicate regularly with their team members and (b) make meaningful and equitable contribution to the final project. Each student could receive different scores for the final project based on their contribution.

You will need to hand in a brief proposal describing the problem and the dataset (max. of two pages). Ideally, the dataset should meet the following requirements:

- at least one continuous (e.g., achievement scores or income)
- at least 6 independent variables (both continuous and categorical variables)
- there is no strict sample size requirement, but more than 100 observations are desirable.

The *project proposal* (max. of two pages) should briefly describe the substantive problem along with a few key references. Since regression analysis is the required

method you should describe the dataset in enough detail such that our TA and I can assess whether the data are suitable for our purposes. In an appendix to the proposal you should give descriptive statistics of the key variables (the dependent and independent variables).

The *project presentation* should provide a summary of your working project paper, based on your project proposal. It should include the following sections: introduction & research questions, data and methods, regression results, and model diagnostics, and if applicable, any concerns you may have. In other words, you will present your ongoing analysis results and discuss any concerns you may have with your peers and teaching staff. The project presentation is a 10-min oral presentation. All the group members are required to present together on the presentation date. Be sure to practice and time yourselves in advance before you give the midterm presentation.

The *final paper* should analyze a formulated research question using appropriate regression techniques. It should cover the following sections: introduction & research questions, data and methods, results, and discussion & conclusions (plus references and appendix). In the "data and methods" section, you should provide descriptive statistics for the important variables that you use in your regression model. Tables and plots should be provided if they are useful in describing the data, but they should be in 1-3 pages. Then, you should try to specify a useful model using substantive theories and model diagnostics. The paper should be written in a style consistent with the major publication outlet in your field (e.g., APA formatting style). More technical descriptions of the regression model (e.g., diagnostics) and SPSS syntax need to be included in an appendix. The maximum number of pages is limited to 12 pages (double spaced; excluding the appendix). The paper should be written as coherently as possible, as if you were submitting it for publication, and should be submitted online. When submitting your final paper, you will also have the opportunity to describe the role each team member played in the final project, including your own role. More details about the final project will be announced on Canvas.

Grading

Courses at Teachers College use the following grading system: A+, A, A-; B+, B, B-; C+, …, F. The symbol W is used when a student officially drops a course before its completion or if the student withdraws from an academic program of the University.

	Requirement	weight for final grade
1.	10 weekly assignments (drop one)	45%
2.	Midterm	20%
3.	Quiz	5%
4.	Proposal & Presentation & Final Paper	30%

If your weighted total points are…	Your final letter grade is…
90-100	A
85-89	A-
80-84	B

75-79	B-
60-74	C
50-59	C-
Below 49	F

Note that A+, B+, and C+ will be determined by the class curve and overall performance in the course.

AI/ChatGPT

Intellectual honesty is vital for an academic community and for the fair evaluation of your work by teaching staff. All work submitted in this course must be your own or that of your group, completed in accordance with the University's academic policies. You should not engage in unauthorized collaboration or make use of ChatGPT or any other AI composition software to complete any of the course assignments.

Services for Students with Disabilities

The College will make reasonable accommodations for persons with documented disabilities. Students are encouraged to contact the Office of Access and Services for Individuals with Disabilities for information about registration (166 Thorndike Hall). Services are available only to students who are registered and submit appropriate documentation." As your instructor, I am happy to discuss specific needs with you as well.

IN Incomplete

The grade of Incomplete is to be assigned only when the course attendance requirement has been met but, for reasons satisfactory to the instructor, the granting of a final grade has been postponed because certain course assignments are outstanding. If the outstanding assignments are completed within one calendar year from the date of the close of term in which the grade of Incomplete was received and a final grade submitted, the final grade will be recorded on the permanent transcript, replacing the grade of Incomplete, with a transcript notation indicating the date that the grade of Incomplete was replaced by a final grade.

If the outstanding work is not completed within one calendar year from the date of the close of term in which the grade of Incomplete was received, the grade will remain as a permanent Incomplete on the transcript. In such instances, if the course is a required course or part of an approved program of study, students will be required to re-enroll in the course including repayment of all tuition and fee charges for the new registration and satisfactorily complete all course requirements. If the required course is not offered in subsequent terms, the student should speak with the faculty advisor or Program Coordinator about their options for fulfilling the degree requirement. Doctoral students with six or more credits with grades of Incomplete included on their program of study will not be allowed to sit for the certification exam.

Email
Teachers College students have the responsibility for activating the Columbia University Network ID (UNI) and a free TC Gmail account. As official communications from the College – e.g., information on graduation, announcements of closing due to severe storm, flu epidemic, transportation disruption, etc. – will be sent to the student's TC Gmail account, students are responsible for either reading email there, or, for utilizing the mail forwarding option to forward mail from their account to an email address which they will monitor.

Religious Holidays
It is the policy of Teachers College to respect its members' observance of their major religious holidays. Students should notify instructors at the beginning of the semester about their wishes to observe holidays on days when class sessions are scheduled. Where academic scheduling conflicts prove unavoidable, no student will be penalized for absence due to religious reasons, and alternative means will be sought for satisfying the academic requirements involved. If a suitable arrangement cannot be worked out between the student and the instructor, students and instructors should consult the appropriate department chair or director. If an additional appeal is needed, it may be taken to the Provost.

Academic Integrity
Students who intentionally submit work either not their own or without clear attribution to the original source, fabricate data or other information, engage in cheating, or misrepresentation of academic records may be subject to charges. Sanctions may include dismissal from the college for violation of the TC principles of academic and professional integrity fundamental to the purpose of the College.

Sexual Harassment and Violence Reporting
Teachers College is committed to maintaining a safe environment for students. Because of this commitment and because of federal and state regulations, we must advise you that if you tell any of your instructors about sexual harassment or gender-based misconduct involving a member of the campus community, your instructor is required to report this information to the Title IX Coordinator, Janice Robinson. She will treat this information as private, but will need to follow up with you and possibly look into the matter. The Ombuds officer for Gender-Based Misconduct is a confidential resource available for students, staff and faculty. "Gender-based misconduct" includes sexual assault, stalking, sexual harassment, dating violence, domestic violence, sexual exploitation, and gender-based harassment. For more information, see http://sexualrespect.columbia.edu/gender-based-misconduct-policy-students

Emergency Plan
TC is prepared for a wide range of emergencies. After declaring an emergency situation, the President/Provost will provide the community with critical information

on procedures and available assistance. If travel to campus is not feasible, instructors will facilitate academic continuity through Canvas and other technologies, if possible.

1. It is the student's responsibility to ensure that they are set to receive email notifications from TC and communications from their instructor at their TC email address.
2. Within the first two sessions for the course, students are expected to review and be prepared to follow the instructions stated in the emergency plan.
3. The plan may consist of downloading or obtaining all available readings for the course or the instructor may provide other instructions.