O'Neill Online SPCN P507 – Spring 2025 (Ver 010525)
Data Analysis and Modeling for Public Affairs

Class Numbers: SPCN 8552/8551
Instructor: Professor Barry Rubin
Cell Phone/Txt: 812-327-3852
Email: rubin@iu.edu
Office Hrs via Zoom by appointment;
  please email, text, or call to set up.
Live Discussion Sessions 6:30-7:30 pm ET
  on 1/13, 2/11, 3/10, 4/7, and 4/28
Zoom: https://iu.zoom.us/j/412226265

Live Labs: Thur, 6:30-8:00 pm ET, on 1/23,
1/30, 2/13, 2/27, 3/13, 4/3, 4/17, and 5/1
TA: Zack Katopodis
TA Email: zkatopod@iu.edu,
TA Phone/Text: (317) 331-1204
TA Office Hrs via Zoom: Thur, 8:00-9:30 pm
  Sun, 4:30-7:00 pm ET; or by appointment
  Zoom: https://iu.zoom.us/j/84348137496

Course Description and Objectives

O'Neill Online P507 provides graduate students with a detailed, intermediate-level perspective on statistical concepts and data science techniques for analyzing and modeling complex systems. The course content includes estimating the parameters of such models based on existing data, testing hypotheses about these systems, and forecasting. As such, the course enables the data analytics process of drawing conclusions based on raw information, and by utilizing historical data in a regression modeling format, provides the framework for both predictive and prescriptive analytics. The context of the course is the application of multivariate regression techniques to problems and policies in public and environmental affairs, the private and non-profit sectors, and society at large.

Multivariate regression analysis is one of the primary tools for statistical modeling for purposes of policy analysis, program evaluation, simulation of systems, data analytics, and general forecasting. Thus, most of the course is devoted to single equation regression models and the extension of these models to a variety of situations. A prerequisite for the class is a graduate-level, introductory statistics course that includes coverage of the simple (two-variable) regression model and an introduction to multivariate regression (e.g., SPCN or SPEA V506, or STAT S520).

The primary learning outcomes for the course are to enable students to:

- explain how regression techniques can be used for data analysis and problem solving;
- apply regression techniques to such problems using the appropriate technology;
- explain the assumptions required to use regression analysis, the impact of violating these assumptions, and how to correct for such violations;
- model and forecast the behavior of systems using regression tools;
- comprehend and critically analyze the results of research employing these tools as presented in academic/professional journals, public/non-profit sector reports, and policy analyses;
- apply several extensions to the basic multivariate regression analysis framework, including logit analysis;
- analyze panel data with fixed and random effects models, and difference-in-differences techniques

O'NEILL
SCHOOL OF PUBLIC AND
ENVIRONMENTAL AFFAIRS

The major topics that will be covered during the semester are:

- the general (multivariate) linear model;
- summary statistics, hypothesis testing, and implicitly linear models;
- multicollinearity, the use of dummy variables, and panel data regression models;
- violations of the general linear model assumptions and errors of specification;
- logit models

Most of the class content is provided via a modular design that has been implemented in the Canvas teaching and learning environment. This includes module content, pre-recorded Tutorials, Live Labs, and five Live Discussion Sessions. Each module is based on a set of class notes, hyperlinked instructional videos that explain specific topics in greater depth, screen capture videos that illustrate several of the major regression techniques and problems, and the accompanying readings in the required textbook. Students are expected to do the readings prior to accessing each module and may find that reviewing the readings again after completing each module is necessary.

The weekly pre-recorded Tutorials have been produced in previous semesters with the instructor and several class members participating. All Lab Sessions are live via Zoom. During these Lab Sessions, which are scheduled for Thursday evenings from 6:30-8:00 pm ET, the TAs will aid with learning R as applied to the course content and provide additional illustrations of concepts, homework assignments, review sessions for exams, and the answers for the homework exercises. In addition, there will be five Live Discussion Sessions delivered via Zoom scheduled from 6:30-7:30 pm ET on Jan 13, Feb 11, Mar 10, Apr 7, and Apr 28. *Both the Tutorials and Labs are a required component of the class. Students are strongly encouraged to attend the live sessions via Zoom. If unable to attend live, students should view the recordings of these sessions which will be posted in the Kaltura Media Gallery.* **Note that the weekly Tutorial should be viewed prior to the respective Lab Session.** Tutorials and the Pre-recorded Lab will be posted to the Kaltura Media Gallery no later than Sunday of the week indicated in the syllabus. Live Labs and recordings of the Live Discussion Sessions will be posted to Kaltura by midnight following the session.

Canvas, email, and Zoom Videoconferencing will be used for student-to-student, student-to-instructor, and student-to-TA interaction. Students can also use the Canvas Discussion forum to post questions, address issues, and assist in establishing groups for homework exercises and the project.

All scheduled office hours for the TAs will utilize Zoom and will be publicly accessible so that multiple students may participate simultaneously. These can be joined at the start of the office hour session or at any time during that period. Office hours by appointment will be private, one-on-one sessions unless a group meeting is specifically requested.

Course Requirements and Grading Criteria

There will be two examinations, each equally weighted. These exams are scheduled for a student-selected 2.5 hour period from March 7-9 and May 5-6**.** Each exam will combine multiple choice questions, short answer essays, and problems. The exams will be comprehensive to the extent that concepts covered in previous sections of the course are required to understand those covered later. The exams will be open book/open note and delivered online. Once a student begins the exam, he/she will have 2.5 hours to complete it. Students are on their honor to not collaborate with anyone else on the exams.

Six homework exercises will be assigned during the course. These will utilize R, RStudio, and R Markdown via students' own computers (or via IUanyWare) to apply the regression techniques covered in the course. Students are required to work collaboratively in teams of two on these homework assignments, using Zoom, phone conferencing, Skype, or other technologies to overcome the impediments of distance. For those students who may be in the same geographic location, face-to-face interaction to complete the exercise is also acceptable. The answers to the assignment will be submitted to the Assignments Area of Canvas by one student in the team on behalf of both students. However, each student should be able to successfully utilize the problem-solving approaches used on the homework assignments. Homework assignments are due on Monday evenings no later than 11:59 pm EST on the dates indicated. Late homework assignments will be accepted with a 10 percent penalty per day late, up to three days, unless an extension is approved by the instructor or TAs prior to the due date. Homework answers must be word-processed and professional quality.

A major project will be required which applies multivariate regression analysis to a data set that students will be responsible for identifying and assembling. Detailed information on the project will be provided mid-March in a separate document in Canvas, and via the indicated Lab. Students are required to work in groups of three on this project, with Zoom and other technologies again utilized to allow for appropriate communications and interaction.

The following grading criteria will be used:

Midterm Exam	20%
Final Exam	20%
Data Set Analysis Project	25%
Homework Exercises	30%
Class Participation	5%

Text and Reading Assignments

The required text for the class is: **Damodar Gujarati and Dawn Porter, 5th ed.,** *Basic Econometrics***, (New York: McGraw-Hill/Irwin, 2009).**

O'NEILL
SCHOOL OF PUBLIC AND
ENVIRONMENTAL AFFAIRS

The text is out of print but may be downloaded as a free PDF file from the P507 Canvas site. Hard copies are also available for purchase online. Additional readings identified in the syllabus will be available via Canvas.

Note that the course syllabus is somewhat flexible and will likely be changed throughout the semester based on pedagogical needs. In addition, this is the first semester in which this course is taught using R, which will also require periodic revision to the course content and/or syllabus.

COVID-19 and Other Issues

As your instructor, I care deeply about the well-being of all students. It is my intent to maintain a positive and safe learning environment in and outside of the classroom. If you are having problems with the class, I encourage you to discuss those with me.

If you have a positive COVID-19 test, have COVID-like symptoms, or have been instructed to quarantine you should not attend class. Attendance for this class is generally required throughout the semester but will never be recorded. For those students who encounter COVID-related issues, alternative arrangements for assignments or make-ups will be offered on a case-by-case basis. Please work with me to determine a path to continue your progress in the class during these absences.

Academic Integrity

All aspects of the IU Student Honor Code (*http://www.iu.edu/~code/*) and the O'Neill SPEA Honor Code (*https://myspea.indiana.edu/doc/ugrad-doc/ugrd-student-honor-code.pdf*) apply to this class, particularly those regarding academic dishonesty and plagiarism.

Note Selling
Several commercial services have approached students regarding selling class notes/study guides to their classmates. Selling the instructor's notes/study guides in this course is not permitted. Violations of this policy will be reported to the Dean of Students as academic misconduct (violation of course rules). Sanctions for academic misconduct may include a failing grade on the assignment for which the notes/study guides are being sold, a reduction in your final course grade, or a failing grade in the course, among other possibilities. Selling a faculty member's notes/study guides individually or on behalf of one of these services using IU email, or via Canvas may also constitute a violation of IU information technology and IU intellectual property policies.

Course Content Ownership
I hold the exclusive right to distribute, modify, post, and reproduce course materials, including all written materials, study guides, lectures, assignments, exercises, and exams. While you are permitted to take notes on the online materials, lectures, and labs for this course for your personal use, you are not permitted to re-post in another forum, distribute, or reproduce content from this course without the express written my permission. Any violation of this course rule will be reported to the appropriate university offices and officials, including to the Dean of Students as academic misconduct.

Disability

If you require assistance or accommodations for a disability, please contact me. You must have established your eligibility for disability support services through the Office of Disabled Student Services in 096 Franklin Hall, 855-7578. For information about support services or accommodations available to students with disabilities, and for the procedures to be followed by students and instructors: http://studentaffairs.iub.edu/dss/.

Sexual Misconduct/Harassment

As your instructor, one of my responsibilities is to help create a safe learning environment on our campus. Title IX and our own Sexual Misconduct policy prohibit sexual misconduct. If you have experienced sexual misconduct, or know someone who has, the University can help.

If you are seeking help and would like to talk to someone confidentially, you can make an appointment with:

    i. The Sexual Assault Crisis Service (SACS) at 812-855-8900
    ii. Counseling and Psychological Services (CAPS) at 812-855-5711
    iii. Confidential Victim Advocates (CVA) at 812-856-2469
    iv. IU Health Center at 812-855-4011

For more information: http://stopsexualviolence.iu.edu/help/index.html. It is also important to know that federal regulations and University policy require me to promptly convey any information about potential sexual misconduct known to me to our campus' Deputy Title IX Coordinator or IU's Title IX Coordinator. In that event, they will work with a small number of others on campus to ensure that appropriate measures are taken, and resources are made available to the student who may have been harmed. Protecting a student's privacy is of utmost concern, and all involved will only share information with those that need to know to ensure the University can assist. Visit http://stopsexualviolence.iu.edu/help/index.html to learn more.

Chat GPT and Other AI Tools

All assignments must be fully prepared by students solely or in homework groups. Any use of generative AI tools (such as ChatGPT) for any part of your work will be a violation of University policy with respect to academic honesty and plagiarism.

O'NEILL ONLINE P507: DATA ANALYSIS AND MODELING
MODULES, TOPICS, READINGS, AND EXERCISES – SPRING 2025[1]

WEEK	MODULE, TOPICS, READINGS, AND EXERCISES

SECTION 1: THE GENERAL (MULTIVARIATE) LINEAR MODEL

Jan 13-19   *Module 1-1: The General Linear Model (GLM)*

Readings: G&P, pp. 34-48, 55-85, 97-102, and 107-134.

Tutorial 1: Introduction to the General Linear Model

Live Discussion Session 1 (Mon, Jan 13, 6:30-7:30 pm): **Introduction to the Class**

Pre-Recorded Lab 1: Accessing R, Intro R Workshop, and Regression in R

Jan 20-26   *Module 1-2: Assumptions and Estimation of the GLM*

Exercise 1 Assigned (Due Mon, Feb 3, by 11:59 pm)

Readings: G&P, pp. 188-196.

Module 1-3: Strength of Relationship and Hypothesis Testing

Readings: G&P, pp. 196-207 and 233-259.

Tutorial 2: Estimation and Strength of Relationship

Live Lab 2 (Thur, Jan 23 6:30-8:00 pm): Intermediate R Workshop

---

**Jan 27-
Feb 2**   *Module 1-4: Implicitly Linear Models and Variable Transformations*

Readings: G&P, pp. 159-173 and 207-213.

**Tutorial 3: Hypothesis Testing in the GLM; Introduction to Implicitly
Linear Models**

**Live Lab 3 (Thur, Jan 30, 6:30-8:00 pm): Hypothesis Testing In R;
Chow Tests in R; Implicitly Linear Regression Overview; Exercise 1 Help Session**

---

[1] All times on the Syllabus are ET.

WEEK	MODULE, TOPICS, READINGS, AND EXERCISES

Feb 3-Feb 9   **Module 1-5: Multicollinearity**

Exercise 1 Due Mon, Feb 3, by 11:59 pm
Exercise 2 Assigned (Due Mon, Feb 17, by 11:59 pm)

Readings: G&P, pp. 320-337.

Tutorial 4: Implicitly Linear Models and Variable Transformations
Intro to Multicollinearity

Prerecorded Lab 4: Implicitly Linear Regression In-Lab Exercise; General R Help;
Exercise 1 Answers Reviewed

Feb 10-16   **Module 1-5: Multicollinearity (continued)**

Readings: G&P, pp. 337-351.
Application Article: Jong Hae Kim, "Multicollinearity and misleading statistical
    results," *Korean Journal of Anesthesiology*,
    2019 Dec; 72(6): 558–569.

Tutorial 5: Multicollinearity

Live Discussion Session 2 (Mon, Feb 10, 6:30-6:30 pm)

Live Lab 5 (Thur, Feb 13, 6:30-8:00 pm): Testing for Multicollinearity;
Exercise 2 Help Session

Feb 17-23   **Module 1-6: Dummy Variables**

Exercise 2 Due Mon, Feb 17, by 11:59 pm
Exercise 3 Assigned (Due Mon, Mar 3, by 11:59 pm)

Readings: G&P, pp. 277-305.

Tutorial 6: Dummy Variables

Pre-Recorded Lab 6: Dummy Variables in R; Exercise 2 Answers Reviewed

---

WEEK	MODULE, TOPICS, READINGS, AND EXERCISES

Feb 24- *Module 1-7: Panel Data - Pooled OLS, Fixed Effects, and Random Effects*
Mar 2   *Models; Difference-In-Differences Models*

Readings: G&P, 591-613.
Application Article: G. Donovan, et al, "The Relationship Between Trees and
Human Health," *American Journal of Preventive Medicine*,
2013, 44(2):139-145.

Tutorial 7: Panel Data

Live Lab 7 (Thur, Feb 27, 6:30-8:00 pm): Panel Data in R;
Exercise 3 Help Session

Mar 3-   *Review and Midterm Exam*
Mar 9

Exercise 3 Due Mon, Mar 3, by 11:59 pm

(No Tutorial this week)

Pre-Recorded Lab: Midterm Review; Exercise 3 Answers Reviewed

---

Mar 7 –   **(Fri, 8:00 am - Sat, 11:59 pm) Midterm Examination via Canvas,**
Mar 9   **2.5 hour time period selected by student**

SECTION 2: VIOLATIONS OF ASSUMPTIONS AND OTHER PROBLEMS IN THE GLM

Mar 10-16 *Module 2-1: Heteroscedasticity*

Data Set Analysis Project Assigned (Due Sat, May 3, by 11:59 pm)

Readings: G&P, pp. 365-401.

Tutorial 9: Introduction to Heteroscedasticity

Live Discussion Session 3 (Mon, Mar 10, 6:30-7:30 pm)

Live Lab 9 (Thur, Mar 13, 6:30-8:00 pm): Midterm Exam Ans Reviewed;
Data Set Analysis Project Discussion and Example

WEEK	MODULE, TOPICS, READINGS, AND EXERCISES

Mar 16-23 *SPRING BREAK*

Mar 24-30 *Module 2-1: Heteroscedasticity (continued)*

Exercise 4 Assigned (Due Mon, April 7, by 11:59 pm)

Readings: G&P, pp. 365-401.

Tutorial 10: Heteroscedasticity

Pre-Recorded Lab 10: Testing and Correcting for Heteroscedasticity

---

Mar 31-Apr. 6 *Module 2-2: Autocorrelation*

Readings: G&P, pp. 412-453.

Tutorial 11: Introduction to Autocorrelation

Live Lab 11 (Thur, Apr 3, 6:30-8:00 pm): Correcting for Heteroscedasticity; Exercise 4 Help Session

---

Apr 7-13 *Module 2-2: Autocorrelation (continued)*

Exercise 4 Due Mon, April 7, by 11:59 pm
Exercise 5 Assigned (Due Mon, April 21, by 11:59 pm)

Readings: G&P, pp. 412-453.

Tutorial 12: Autocorrelation

Live Discussion Session 4 (Mon, Apr 7, 6:30-6:30 pm)

Pre-Recorded Lab 12: Testing and Correcting for Autocorrelation; Exercise 4 Answers Reviewed

---

Apr 14-20 **Module 2-3: Errors of Specification and Proxy Variables**

Readings: G&P, pp. 467-486.
Application Article: Barry Rubin, Specification of Wage Determination
Mechanisms in Urban Labor Markets, *Socio-Economic Planning Sciences*, 1985, 19(6):387-398.

Tutorial 13: Specification Errors

Live Lab 13 (Thur, Apr 17, 6:30-8:00 pm): Data Analysis Project Help Session; Exercise 5 Help Session

SECTION 3: ALTERNATIVE FORMS AND EXTENSIONS OF REGRESSION ANALYSIS

Apr 21-
Apr 27   **Module 3-1: The Linear Probability and Logit Models**

Exercise 5 Due Monday, April 21, by 11:59 pm
Exercise 6 Assigned (Due Monday, April 28, by 11:59 pm)

Readings: G&P, pp. 552-566.
Denvil Duncan, Venkata Nadella, et al, "The Road Mileage User-Fee," *Transport Policy,* 2017, 53:70-78.

Tutorial 14: Logit Models

Pre-Recorded Lab 14: R Examples of Logit Models; Exercise 6 Help Session; Exercise 5 Answers Reviewed

Apr 28- **Module 3-2: Distributed Lag and Autoregressive Models; Research Example.**
May 4 **and Final Exam Review**

Exercise 6 Due Mon, April 28, by 11:59 pm
Data Analysis Project Report and Exec Summary Due Sat, May 3, by 11:59 pm

Tutorial 15: Distributed Lag and Autoregressive Models; Research Example

Live Discussion Session 5 (Mon, Apr 28, 6:30-7:30 pm)

Live Lab 15 (Thur, May 1, 6:30-8:00 pm): Project Help Session; Exercise 6 Answers Reviewed

Live Final Exam Review (Sun, May 4, 6:30-6:30 pm)

May 5-6 *(Mon-Tues)* **Final Examination via Canvas, 2.5 hour time period selected by student**