PORTLAND STATE UNIVERSITY

USP 654: Data Analysis II

Liming Wang	**Fall 2023**
Office: URBN 350D	**M 4:00-6:30 FMH B157**
Office hours: Tue 11am-1pm	**M 6:40-7:30 FMH B157**
or by appointment	**4 credits**
Email: lmwang@pdx.edu	**Course website: Canvas**
Zoom link: https://pdx.zoom.us/j/5037255130	

1. **Description and Objectives**

USP 654 provides instruction in applied linear regression modeling, a powerful technique and perhaps the most widely used advanced statistical method in the social sciences, covering dummy variables, interactions, nonlinear relationships and diagnostics. In addition, the course will offer a quick introduction to modeling non-continuous data--including count and discrete data - also common in the social sciences. Students will leave this course confident in their abilities to begin their own statistical investigations and ready to contribute to empirical knowledge in their chosen fields of study.

This course is in applied statistics, and is not a substitute for more rigorous courses in mathematics for those students interested in understanding the underlying statistical theory. Our focus here is on applying and critically evaluating model applications in a way that respects the theory behind them, while taking mathematical derivations as given. We can learn to use a tool skillfully and as its designers intended without a complete understanding of how it was made. In practice, modeling human behavior is often as much art as science.

By the end of the class, students are expected to:

1. Develop a solid understanding of fundamental concepts of regression analysis.
2. Acquire proficiency in using statistical software and tools such as R and SPSS, as well as relevant libraries and packages, for regression analysis.
3. Interpret and critically evaluate results of regression analysis of their own and others: Students should develop the ability to assess and critique regression analysis results, as well as the potential biases and confounding factors that may affect their conclusions.

4. Design and execute projects utilizing regression analysis as the primary method of analysis: Students should be able to plan and carry out data analysis projects from start to finish, including formulating research questions, collecting and processing data, selecting appropriate methods, and presenting their results.
5. Communicate findings effectively to diverse audiences: Students should be able to explain their results in a clear and engaging manner, both in writing and through oral presentations, adapting their communication style to suit different audiences.

1. **Prerequisite:** USP 634 Data Analysis I or another introductory course in Statistics
2. **Texts and Readings**

Required Text:

- Gordon, Rachel (2015), Regression Analysis for the Social Sciences (2nd edition), Routledge.
- Grolemund and Wickham (2023), R for Data Science (2nd eidtion), O'Reilly. (Available online at https://r4ds.hadley.nz/)

Additional Readings: Various journal articles and other readings will be assigned throughout the term to support the main texts. Some will be in connection to lecture and others specific to homework problems. All readings will be shared in advance on the course website, in part because I want the flexibility to adjust readings based on class interests (or suggestions).

Other Texts I recommend, for now or in the future:

Cohen, J., Cohen, P., West, S.G., & Aiken, L.S. (2003). Applied multiple regression/correlation analysis in the behavioral sciences (Third Edition). Mahwah, NJ: Erlbaum. ISBN: 0-8058-2223-2.

Alan, A., & Barbara, F. (1986). Statistical methods for the social sciences. Dellen Publishing Company, San Francisco and Collier Macmillan Publishers, London, 4(1), 105-132.

Gujarati, D. N. (2009). Basic econometrics. Tata McGraw-Hill Education.

Hensher, D. A., Rose, J. M., & Greene, W. H. (2005). Applied Choice Analysis: a Primer. Cambridge University Press.

Ben-Akiva, M. E., & Lerman, S. R. (1985). Discrete Choice Analysis: Theory and Application to Travel Demand (Vol. 9). MIT press.

Grading

The course requirements include assigned readings, three assignments, DataCamp exercises, and a research project (paper and presentation).

1. **Homework (45%)**

Over the course of the term, you will complete 3 homework assignments and a class project based on readings, lectures, and lab activities. The homework will consist of a mix of theoretical, empirical, and critical interpretation of analysis done by others. You are welcome to consult with your classmates, but everyone is responsible for handing in their original work, in their own words.

2. DataCamp Courses (20%): We will use the interactive courses on DataCamp to learn to use R for regression analysis. We will dedicate some of the lab hours for you to work on these courses, but you're expected to finish them outside labs. There are no deadlines for these courses as long as they are completed by 12/04. Use this link to register your account to get access to all DataCamp courses free of charge:

- Introduction to Regression in R
- Inference for Linear Regression in R
- Intermediate Regression in R
- Generalized Linear Models in R

3. Research Project (35% = presentation 5% + paper 30%): In this class project, you will explore a topic of your choosing in depth, applying regression analysis covered in this class (and beyond) for your data analysis. Ideally, you would build upon work you have already done for other courses, such as USP 634 DA I, USP 630 Research Design. The paper must include original data analysis. This most likely will be analysis of secondary data (check out the Tools and Datasets section of the course website for some potential datasets for your class project). Details:

- 20 pages, double-spaced, not including figures, tables and references;
- Initial proposal due on 10/09. It should include a brief description of the topic, a list of key references (including academic literature), specific research question(s) and hypotheses, and the data source(s) you plan to use;
- Project updates due on 11/13. This should include a draft of the introduction/background sections, a brief literature review, the methodology and description and descriptive statistics of your data, as well as initial regression results;
- Project presentation on 12/04 (tentative);
- Final paper due on 12/08.

1. **Grading**

Homework assignments and class project will make up your entire grade in the course. I use standard grading cutoffs (A=90%, B=80%, C=70%, D=60%, F<60%).

Late assignments will be marked down – one-third of a grade per day late. "One-third of a grade" is, for example, from A to A-, B+ to B, etc. (or 3.3 points using the 100 point scale). We generally do not allow students to turn things in late without assessing this penalty, except in unusual circumstances, e.g. medical emergencies. A request for such an extension must be made *before* the due date. Having too much work in other classes or at work/internship does not count. All students have those challenges.

1. **SCHEDULE (subject to adjustment):**

Date	Week	Topic	Reading/HW
09/25	1	No class Introduction;	Chapter 1-2
10/02	2	Bivariate Regression	Chapter 5 HW1 handout
10/09	3	Multiple Regression	Chapter 6 **Project proposal due**
10/16	4	Dummy Variables	Chapter 7 **HW1 due** HW2 handout
10/23	5	Interactions	Chapter 8
10/30	6	Nonlinear Relationships	Chapter 9 Chapter 10
11/06	7	Indirect Effects; Omitted Variable and other Biases	**HW2 due** HW3 handout
11/13	8	Model Diagnostics: Outliers, Heteroskedasticity, and Multicollinearity	Chapter 11 **Project update due**

11/20 9	Causal Inference	Chapter 12
11/27 10	Generalized Linear Models; Where to go from here	**HW3 due**
12/04 11		**Project presentation**
12/08 11		**Final paper due**

1. **Notes**

1. If my office hours conflict with your schedule, send me a Google calendar invite for an appointment. Follow this instruction if you don't know how to use Google calendar.
2. Submit in your homework assignments to Canvas/Assignments.
3. Extensions on homework due dates will only be granted in exceptional circumstances and only if I am contacted before the due date.
4. Academic dishonesty will result in a score of zero for the assignment in question.
5. Please be aware of the University's policies on withdrawals and incompletes:
    - Withdrawals: https://www.pdx.edu/registration/withdraw-cancellation-and-drops
    - Incompletes: http://pdx.smartcatalogiq.com/en/2016-2017/Bulletin/Graduate-Studies/Enrollment/Incompletes
6. Good advice paraphrased from a former instructor of this course: Be active! The more you do in this course, the more you will learn. Apply the tools and concepts to your own/other data; see what you find; if it is interesting/confusing/etc., bring & show it to me! I can get you started, but good data analysts are driven by a curiosity they discover in themselves.

1. **ACCESS AND INCLUSION FOR STUDENTS WITH DISABILITIES**

PSU values diversity and inclusion; My goal is to create a learning environment that is accessible, equitable,, inclusive, and welcoming. I am committed to fostering mutual respect and full participation for all students. If any aspects of instruction or course design result in barriers to your inclusion or learning, please notify me. Additionally, the Disability Resource Center (DRC) provides reasonable accommodations for students who encounter barriers in the learning environment. The DRC works with students who have physical, learning, cognitive, mental health, sensory, chronic illness, and other disabilities.

If you have, or think you may have, a disability that may affect your work in this class and feel you need accommodations, contact the Disability Resource Center to schedule an appointment and initiate a conversation about reasonable accommodations.

If you already have accommodations, please contact me to make sure that I have received your DRC Faculty Notification Email so we can discuss your accommodations.

The DRC is located in 116 Smith Memorial Student Union, Suite 116. You can also contact the DRC at 503-725-4150 or, drc@pdx.edu. Visit the DRC online at https://www.pdx.edu/disability-resource-center.

1. **Title IX reporting obligations**

Portland State is committed to providing an environment free of all forms of prohibited discrimination and sexual harassment (sexual assault, domestic and dating violence, and gender or sex-based harassment and stalking). If you have experienced any form of gender or sex-based discrimination or sexual harassment, know that help and support are available. Information about PSU's support services on campus, including confidential services and reporting options, can be found on PSU's Sexual Misconduct Prevention and Response website at: http://www.pdx.edu/sexual-assault/get-help or you may call a confidential IPV Advocate at 503-725-5672 or schedule Online at https://psuwrc.youcanbook.me . You may report any incident of discrimination or discriminatory harassment, including sexual harassment, to:

- PSU's Title IX Coordinator: Julie Caron by calling 503-725-4410, via email at titleixcoordinator@pdx.edu or in person at Richard and Maureen Neuberger Center (RMNC), 1600 SW 4th Ave, Suite 830
- Deputy Title IX Coordinator: Yesenia Gutierrez by calling 503-725-4413, via email at yesenia.gutierrez.gdi@pdx.edu or in person at RMNC, 1600 SW 4th Ave, Suite 830

Please be aware that all PSU faculty members and instructors are required to report information of an incident that may constitute prohibited discrimination, including sexual harassment and sexual violence. This means that if you tell me about a situation of sexual harassment or sexual violence that may have violated university policy or student code of conduct, I have to share the information with my supervisor, the University's Title IX Coordinator or the Office of the Dean of Student Life. However, the Title IX Coordinators will keep the information confidential and refer you to a confidential advocate. For more information about Title IX please complete the required student module *Creating a Safe Campus* in your Canvas.