

## JSGS 803 - QUANTITATIVE METHODS AND RESEARCH DESIGN

	UNIVERSITY OF SASKATCHEWAN CAMPUS	UNIVERSITY OF REGINA CAMPUS
INSTRUCTOR:	Haizhen Mou	
PHONE:		
E-MAIL:	<a href="mailto:Haizhen.mou@usask.ca">Haizhen.mou@usask.ca</a>	
OFFICE HOURS:	By appointment	
OFFICE LOCATION:	DIEF 158	
TERM/SEMESTER:	Fall 2022	
ROOM:	Online, synchronous	
DATE AND TIME:	Tuesdays, 1-4pm (Sept. 6 – Dec. 6)	

The syllabus for this course is comprised of this document and a companion document titled “JSGS Common Syllabus 2022-23.”

### CALENDAR DESCRIPTION

Provides students with the statistical concepts and techniques required for conducting research and critically evaluating empirical studies. Topics include statistical inference, sampling theory, and data and regression analysis as applied to problems in public policy.

### LEARNING OBJECTIVES

JSGS programs, courses, and assignments are designed to prepare students for successful careers. Student performance is assessed based on a competency framework shaped through ongoing collaboration with researchers, alumni, and other public sector partners. The specific readings, assignments and activities in JSGS 803 will help you acquire and demonstrate competence in the following:

- **Evidence and Strategic Thinking** – develops and applies an evidence-informed approach to policy issues and policy options;

## **COURSE CONTENT AND APPROACH**

This course is designed to introduce you to the basic principles of bivariate and multivariate regression analysis and selected advanced models, and to apply the regression model to problems in public policy research. Various modifications to the regression model will be examined, as will several problems that often plague regression models. Finally, you will become familiar with the regression routines of STATA. This software will be used in completing class assignments.

This course emphasizes the application of statistical procedures more than the theoretical or mathematical principles behind them. While you will learn the basic theoretical principles of regression analysis, the overall objective of the course is to learn how to apply this method to diverse empirical problems in public policy.

The course has a lecture format. However, discussion is encouraged.

## **COURSE FORMAT**

### *Online, Synchronous Instruction*

This course is designed for online participation in real time (also known as synchronous instruction). Scheduled classes complement learning material found on PAWS. Please review the learning material in advance of each scheduled class. Your regular attendance in class offers the best opportunity to ask questions, make connections, and participate in discussions based on the learning material.

## **REQUIRED READINGS**

Leo H. Kahane, *Regression Basics* 2nd edition (Thousands Oaks: Sage, 2008). ISBN: 9781412951265.

Supplementary reading if you need more detailed statistics knowledge:

John Fox, *Applied Regression Analysis and Generalized Linear Models*, Third Edition (April 15, 2015). SAGE Publications, Inc. ISBN-13: 978-1452205663. A R Companion to this book is at <https://socialsciences.mcmaster.ca/jfox/Books/Companion/index.html>

## **COURSE OUTLINE (SUBJECT TO CHANGES DURING THE TERM)**

### **Session 1 – Introduction, correlation and linear regression (September 6)**

#### **REQUIRED READINGS**

- > Kahane, *Regression Basics*, Chapter 1; Chapter 2, pp. 19-29.
- > Michael S. Lewis-Beck, *Applied Regression: An Introduction*, Chapter 1, pp. 9-26. (<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412983440.n1>)

#### **SUPPLEMENTARY READING**

- > Levin, Jack and James Alan Fox, *Elementary Statistics in Social Research* (New York: Harper & Row, 2004) Chapters 7-9 “From Description to Decision Making”, pp. 128-192. For a review of basic statistical concepts, see Chapter 3-6.

> John Richard, Jennifer Hove, and Kemi Afolabi, "Understanding the Aboriginal/Non-Aboriginal Gap in Student Performance" C.D. Howe Institute Commentary, No. 276 (December 2008), available online at [http://www.cdhowe.org/pdf/commentary\\_276.pdf](http://www.cdhowe.org/pdf/commentary_276.pdf).

## **Session 2 – Hypothesis testing and model performance (September 13)**

### REQUIRED READINGS

- > Kahane, Chapter 3
- > Michael S. Lewis-Beck, Applied Regression: An Introduction, Chapter 2, pp. 26-43. (<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412983440>)
- > Edward Tuft, Data Analysis for Politics and Policy (Englewood Cliffs, N.J.: Prentice-Hall, 1974), Chapter 3, "Example 5: Comparing the Slope and the Correlation Coefficient", pp. 101-107.
- > Christopher H. Achen, Interpreting and Using Regression, Chapter 4, "Comparing Substantive and Statistical Significance", pp. 46-51. (<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412984560.n4>)

### SUPPLEMENTARY READING

- > Joseph P. Newhouse, "Medical-Care Expenditure: A Cross-National Survey," Journal of Human Resources, 12:1 (Winter, 1977), pp.115-125.

## **Session 3 – Regression model assumptions and the analysis of residuals (September 20)**

### REQUIRED READINGS

- > Kahane, Chapter 2, pp. 30-34; Chapter 7.

## **Session 4 – The multiple regression model: partial slopes (September 27, A1 is due)**

### REQUIRED READINGS

- > Kahane, Chapter 4.
- > Michael S. Lewis-Beck, Applied Regression: An Introduction, Chapter 3, pp. 47-54. <http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412983440.n3>

### SUPPLEMENTARY READINGS

- > Stuart Landon, Melville L. McMillan, Vijay Muralidharan and Mark Parsons, "Does Health-Care Spending Crowd Out Other Provincial Expenditures?" Canadian Public Policy 32:2 (June, 2006), pp. 121-142. <http://library.usask.ca/scripts/remote?URL=http://www.jstor.org/stable/4128724>

## **Session 5 – The multiple regression model: dummy variables and interaction effects (October 4)**

### REQUIRED READINGS

- > Kahane, Chapter 5, pp. 86-101.
- > Achen, Chapter 2, pp. 12-29. <http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412984560.n2>

### SUPPLEMENTARY READINGS

- > Mohammed H.I. Dore and Surendra Kulshreshtha, "The Labor Market and Rural-Urban Differences among First Nations: The Case of Saskatchewan," Journal of Socio-Economics, 32:2 (May, 2003), pp. 147-159.

> Card, David and Krueger, Alan B, 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *American Economic Review*, 84:4 (September, 1994), pp. 772-793.

<https://www.jstor.org/stable/pdf/2118030.pdf?refreqid=excelsior%3A716af50c2dbe7f10ac3003949973ed14>

### **Session 6 – The multiple regression model: choosing a specification – including Outliers and data transformations (October 11)**

#### REQUIRED READINGS

> Kahane, Chapter 5, pp. 79-83; Chapter 7.

> Edward Tufte, *Data Analysis for Politics and Policy* (Englewood Cliffs, N.J.: Prentice-Hall, 1974), Chapter 3, "Example 6: Interpretation of Regression Coefficients when the Variables are Re-expressed as Logarithms (with Five Examples)", pp. 108-134.

> Achen, Chapter 5, "Choosing a Specification", pp. 51-77.

<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412984560.n5>

#### SUPPLEMENTARY READINGS

> Robert W. Jackman, "The Politics of Economic Growth in the Industrial Democracies, 1974-80: Leftist Strength or North Sea Oil?," *The Journal of Politics*, 49: 1 (Feb., 1987), pp. 242-256.

<http://library.usask.ca/scripts/remote?URL=http://www.jstor.org/stable/2131143>

> John Fox, *Regression Diagnostics* (Newbury Park, CA: Sage, 1991), Chapter 4 "Outlying and Influential Data", pp. 21-40.

<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412985604.n4>

> Schroeder et al., *Understanding Regression Analysis: An Introductory Guide*, Chapter 5, "Problems and Issues of Linear Regression", pp. 65-80.

<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412986410.n5>

### **Session 7 – Presentation and midterm review (October 18, A2 is due)**

### **Session 8 – logistic regression and time series data (October 25)**

#### REQUIRED READINGS

> Logistic: Kahane, Chapter 5, pp.83-86.

> Time series data: Kahane, Chapter 6; Chapter 7, pp. 132-138.

> Fred C. Pampel, *Logistic Regression: A Primer* (Newbury Park, CA: Sage, 2000), Chapter 1 "The Logic of Logistic Regression"

(<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/10.4135/9781412984805.n1>) and Chapter 2 "Interpreting Logistic Regression Coefficients", pp. 1-30.

<http://library.usask.ca/scripts/remote?URL=http://dx.doi.org/DOI:10.4135/9781412984805>

#### SUPPLEMENTARY READING

> Ross Finnie, "Who Moves? A Logit Model Analysis of Inter-Provincial Migration in Canada," *Applied Economics*, 36:16 (2014), pp. 1759-1779. <https://doi.org/10.1080/0003684042000191147>

> Vishnu Kapur and Kisalaya Basu, "Drug Coverage in Canada: Who Is at Risk?" *Health Policy* 71 (2005), 181-193

[https://www.researchgate.net/publication/8123417\\_Drug\\_coverage\\_in\\_Canada\\_Who\\_is\\_at\\_risk](https://www.researchgate.net/publication/8123417_Drug_coverage_in_Canada_Who_is_at_risk)

- > Haizhen Mou, Michael M. Atkinson, and Ata-Ul Munim, "The Cost of Government: Decomposing Provincial Expenditures, 1981 – 2007," *Canadian Public Policy* 40:1 (March, 2014), pp. 84-97.  
[https://www.jstor.org/stable/24365079?seq=1#metadata\\_info\\_tab\\_contents](https://www.jstor.org/stable/24365079?seq=1#metadata_info_tab_contents)

### **Session 9 – Hierarchical Linear Modelling (November 1)**

#### REQUIRED READINGS

- > Raudenbush, Stephen, and Anthony S. Bryk. "A Hierarchical Model for Studying School Effects." *Sociology of Education* 59, no. 1 (1986): 1-17. doi:10.2307/2112482.
- > Gelman, Andrew. "Multilevel (Hierarchical) Modeling: What It Can and Cannot Do." *Technometrics* 48, no. 3 (2006): 432-35. <http://www.jstor.org/stable/25471214>.

#### SUPPLEMENTARY READING

- > Haizhen Mou and Michael M. Atkinson, "Want to Improve Math Scores? An Empirical Assessment of Recent Policy Interventions in Canada" *Canadian Public Policy* 46:1 (2020), pp. 107–124.  
<https://www.utpjournals.press/doi/abs/10.3138/cpp.2019-025?journalCode=cpp>

### **November 8: Reading week – no class**

### **Session 10 – Introduction to Causal Modelling (November 15, A3 is due)**

#### REQUIRED READINGS

- > Lechner, Michael. The estimation of causal effects by difference-in-difference methods. 2011.  
[https://michael-lechner.eu/ml\\_pdf/journals/2011\\_Lechner\\_DiD\\_2011\\_ECO%200403%20Lechner\\_darf%20aufs%20Netz.pdf](https://michael-lechner.eu/ml_pdf/journals/2011_Lechner_DiD_2011_ECO%200403%20Lechner_darf%20aufs%20Netz.pdf)
- > Luellen, Jason K., William R. Shadish, and M. H. Clark. "Propensity scores: An introduction and experimental test." *Evaluation Review* 29:6(2005), pp. 530-558.  
[https://www.researchgate.net/publication/7522646\\_Propensity\\_Scores\\_An\\_Introduction\\_and\\_Experimental\\_Test](https://www.researchgate.net/publication/7522646_Propensity_Scores_An_Introduction_and_Experimental_Test)
- > Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. "Identification and estimation of treatment effects with a regression-discontinuity design." *Econometrica* 69:1 (2001), pp. 201-209.  
<http://www.jstor.org/stable/2692190>

#### SUPPLEMENTARY READING

- > Howe, Katherine B., Christian Suharlim, Peter Ueda, Daniel Howe, Ichiro Kawachi, and Eric B. Rimm. "Gotta catch'em all! Pokémon GO and physical activity among young adults: difference in differences study." *bmj* 355 (2016). <https://www.bmj.com/content/355/bmj.i6270.full>
- > Mendola, Mariapia. "Agricultural technology adoption and poverty reduction: A propensity-score matching analysis for rural Bangladesh." *Food policy* 32, no. 3 (2007): 372-393.  
<https://www.sciencedirect.com/science/article/abs/pii/S0306919206000790>
- > Barber, Andrew and West, Jeremy, "Conditional Cash Lotteries Increase COVID-19 Vaccination Rates" (July 26, 2021). Available at SSRN: <https://ssrn.com/abstract=3894034>
- > Abou-Chadi, T., & Krause, W.. The Causal Effect of Radical Right Success on Mainstream Parties' Policy Positions: A Regression Discontinuity Approach. *British Journal of Political Science*, 50:3(2020), 829-847. <https://doi.org/10.1017/S0007123418000029>

## **Session 11 – Introduction to Machine Learning Methods (November 22, to be confirmed)**

### REQUIRED READINGS

> Sutton, Oliver. "Introduction to k nearest neighbour classification and condensed nearest neighbour data reduction." *University lectures, University of Leicester* (2012): 1-10.  
<https://staff.fmi.uvt.ro/~daniela.zaharie/dm2018/ro/TemeProiecte/Biblio/kNN/CondensedNearestNeighbor.pdf>

### SUPPLEMENTARY READING

> Richardson, Adam and Mulder, Thomas and Vehbi, Tugru, "Nowcasting New Zealand GDP Using Machine Learning Algorithms" (September 28, 2018). CAMA Working Paper No. 47/2018, Available at <http://dx.doi.org/10.2139/ssrn.3256578>

> J. Ramteke, S. Shah, D. Godhia and A. Shaikh, "Election Result Prediction Using Twitter Sentiment Analysis," 2016 International Conference on Inventive Computation Technologies (ICICT), 2016, pp. 1-5, doi: [10.1109/INVENTIVE.2016.7823280](https://doi.org/10.1109/INVENTIVE.2016.7823280)

## **Session 12 – Term project presentation and discussion (Nov. 29, A4 is due)**

## **Session 13 – Term project presentation and discussion (Dec. 6)**

## **EVALUATION**

There will be four assignments. Please upload assignments on the course webpage before the deadlines.

Assignment 1 Bivariate Regression	Due at 9am on Sept. 27	20%
Assignment 2 Multiple Regression	Due at 9am on Oct. 18	20%
Assignment 3 Logistic Regression	Due at 9am on Nov. 15	30%
Assignment 4 Term Project	Due at 9am on Nov. 29	30%

In this course, I assume that most of you have no experience in using STATA. I will be explaining how to use STATA for particular problems. Should you desire more exhaustive resources, there are many manuals available online (e.g. at <http://www.stata.com/links/resources1.html>).

While the STATA software is expensive you can get access to STATA for free by following the instructions on the website of "ITS Service and Support", at <https://wiki.usask.ca/display/public/ISDKB/STATA>. The university also offers a virtual lab access to the software at <https://vlab.usask.ca/EricomXml/accessportal/start.html#/login>. If you need help, ask the ITS support staff.

## **ENROLLMENT**

Class enrollment will be normally limited to 30 students.

## **INTELLECTUAL PROPERTY ACKNOWLEDGEMENT**

This course was developed by Haizhen Mou.