

Department of Sociology

Winter Term 2018

Advanced Multivariate Statistical Analysis

Sociology 9007B

Wednesdays 1:30 – 4:20pm

Room TBA (Labs in SSC 1000)

DRAFT AUGUST 2017: SUBJECT TO CHANGE

Course and Instructor Information

Professor: Rachel Margolis, Ph.D.

Office: Social Science Center 5326

Email: rachel.margolis@uwo.ca

Office Hours: Mondays 12pm-1:30pm

3 seminar hours, 0.5 course

Course Description

In this course we will cover the most common statistical techniques in the practice of sociology – linear regression, logistic regression, survival analysis (event history analysis), and fixed effects models. We will discuss the uses of these techniques and the assumptions that we make when using them. We will also spend time discussing how to interpret results and how to choose the best method for the research question. We will cover issues that we need to pay attention to, such as nonlinearity, outliers, multicollinearity, and interaction terms. We will also discuss basics of panel data analysis. Throughout the course, we will discuss how to develop an answerable research question, how to choose the best modeling strategy for that question, and how to interpret the results of quantitative analysis in light of relevant hypotheses.

We will have an applied portion of the class held in the computer lab (SSC 1000). We will conduct data analysis using STATA, a statistical analysis software package. In the lab, we will talk through basic issues that come up such as dealing with missing data; saving data, code, and output; and making tables.

The last portion of the course will focus on writing about multivariate analysis. Once you've done some analysis, it's important to be able to communicate clearly what you did, why, and what you found. You must come into the course with a project in mind, for which you use one of the methods that we learn about in class. The project will be the majority of a paper including the research questions, contributions, data, methods, results and discussion sections. In workshop format, students will present their projects to the class.

Learning Outcomes

1. Understand the assumptions made when using linear regression, logistic regression and survival analysis.
2. Interpretation of key results from multivariate models.
3. Conduct analysis with these methods using Stata.
4. Critically evaluate published research utilizing common multivariate methods in sociology.
5. Communicate about the results of multivariate analysis in presentation form and in writing.

Prerequisites

Students should have taken an introductory statistics course at the graduate level that covers descriptive statistics, probability, bivariate regression, ANOVA, and linear regression. There will be a diagnostic test in week 1. Students who do not have the prerequisite knowledge will be referred to an introductory course before they can take this course.

We will be using Stata, a data analysis software, in class. Students should have been exposed to this in introductory statistics (Soc 9001A). Those with little computer literacy are strongly advised to make use of the opportunities provided by the Social Science Computing Lab with non-credit courses and/or consulting desk services.

We will be doing data analysis in the SSC computer lab in room 1000. Students should have a computer account with the SSNDS network and should make sure their login works prior to the first computer lab session. In order to have access to STATA in the computer lab, students must:

- 1) Have active Western accounts
- 2) Have subscribed to “Western Identity Manager” and synchronized their passwords through the “Profile” tab.
- 3) If students are having problems with any of the above, then they can either call the SSNDS main office at 519-661-2152 or call the ITS Help Desk at 519-661-3800.

Requirements and Evaluation

Grading will be based on a midterm test, quiz, and a data analysis project.

Midterm Test (35%), Feb 14, 2018

Quiz (10%), March 14, 2018

Data Analysis Project

Research Questions and Data Sections (10%), Due Feb 28, 2018

Presentation (20%), in class April 4, April 11, 2018

Final Version Due April 23, 2018 at 10am (25%)

Missed Exams and Late Assignments: In case of medical illness or other circumstance and an exam cannot be taken, students can bring documentation in writing. Late assignments without documentation are marked down 5 points (of 100) per day (24-hour period) late. Anything submitted after the time/date due is one day late.

The Data Analysis Project (More instructions at end of course outline)

The data analysis project is designed to help students develop a research question, conduct data analysis, make tables, and write and present the method and results. Throughout the semester, you will gain skills that you will then use in your research project. The project can be on any topic, but it must use one of the methods that we are covering in the course. *Depending on the class size, I may decide to have students do the project in pairs.*

There are 3 graded components to the project:

1) Research Questions and Data: (1-2 pages) Students will submit their research question and data section for the paper on February 28 in class. Include here a section on the data that you will use to answer the research questions. Explain why the data are well suited to the research questions.

2) In-class presentation: During the last few weeks of the term, students will present their projects to the class. The in-class presentations will take place on April 4 and April 11. Students should prepare a 10-13 minute presentation on their research project.

3) Research papers are due April 23, 10am. Students will turn in the following sections of a research paper: Research Questions, Contributions, Data, Methods, Results, Tables, and References. Directions for formatting can be found at the end of the course outline.

Overview of Course Schedule

| Week | Date | Topic | Notes Whether We Have a Lab Session 1:30-2:30pm or 3:30-4:30pm |
|---|----------|--|--|
| 1 | Jan 10 | Introduction, Short Review, Diagnostic Test, Working with Data | |
| 2 | Jan 17 | Linear Regression Review | Tentative |
| 3 | Jan 24 | Linear regression- making a model, interactions | Tentative |
| 4 | Jan 31 | Regression Diagnostics/Logit models I | |
| 5 | Feb 7 | Logit models II | Tentative |
| 6 | Feb 14 | Midterm Test (In Class) | |
| | Feb 21 | Reading Week- No class | |
| 7 | Feb 28 | Survival Analysis I | |
| 8 | March 7 | Survival Analysis II | Tentative |
| 9 | March 14 | Quiz, Causality, Instrumental Variables | |
| 10 | March 21 | Fixed Effects Models | |
| 11 | March 28 | Advanced Topics | |
| 12 | April 4 | Workshop Student Projects | |
| 13 | April 11 | Workshop Student Projects | |
| April 23, 2018 Research papers are due at 10am. | | | |

* Note: If there is no lab session, then class will be held in our regular classroom from 1:30 – 4:30.

Required Text

There are no required texts for this course. Recommended texts are below. Other readings will be provided on the owl/sakai course website.

Recommended Texts (These texts will be on course reserve at Weldon Library)

- Allison, Paul D. 1999. *Multiple Regression: A Primer*. Pine Forge Press.
This book is a friendly book which clearly describes the purposes of linear regression, the assumptions made, and how multiple regression differs from other statistical techniques.
- Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate Statistics*. University of Chicago Press.
This book is a very good resource for writing about the results of linear and logistic regression, and survival analysis. Also consult this book when making tables.
- Gordon, Rachel A. 2010. *Regression Analysis for the Social Sciences*. Routledge Press.
This text is an applied text on linear regression which focuses on guiding students through using data, running, and interpreting regressions. It includes examples for commands using STATA and SAS.
- Pampel, Fred C. 2000. *Logistic Regression: A Primer*. Thousand Oaks, California. Sage Publications.
- Cleves, M., R.G. Gutierrez, W. Gould, and Y.V. Marchenko. 2010. *An Introduction to Survival Analysis Using Stata, Third Edition*. Stata press.
This is a helpful text for those who are using survival analysis for a project. Not necessary to purchase for the course.

For more information on specific methods, consult alternative texts. Below are a few recommended texts.

- Agresti, A. (2007) *An Introduction to Categorical Data Analysis*, Wiley.
- Agresti, A. and B. Finlay (1999) *Statistical Methods for the Social Sciences*, Third edition, Prentice Hall. (the fourth edition should be available early in 2009)
- Fox, J. (1997) *Linear Statistical Models and Related Methods*, Wiley.
- Blossfeld, H.P., K. Golsch, and G. Rohwer (2007) *Event History Analysis with Stata*, Lawrence Erlbaum Associates.
- Hosmer, D.W. & S. Lemeshow (2000) *Applied Logistic Regression*, Wiley.
- Powers, D.A. and Y. Xie (2000) *Statistical Methods for Categorical Data Analysis*, Associated Press.
- Singer, J.D., and Willett, J.B. (2003). *Applied Longitudinal Data Analysis*, Oxford University Press.
- Stevens, J. (1996) *Applied Multivariate Statistics for the Social Sciences*, Third edition. New Jersey, Lawrence Elbaum.

Readings

Methods readings almost all come from the 5 recommended texts which can be bought at the bookstore or borrowed from Weldon course reserve. All example readings and a few methods readings are journal articles and will be available on the course website or through the library website.

Week 1: Introduction

Methods reading: Review Gordon textbook Chapters 1-4 for a refresher on linear regression.

Methods reading: Start reading Allison's primer "Multiple Regression"

Example reading: Burgard and Ailshire. 2013. Gender and Time for Sleep among U.S. Adults. *American Sociological Review* 78:51.

Week 2: Linear Regression

Methods reading: Miller Chapters 5 and 6

Methods reading: Gordon textbook Chapters 5-6

Example reading: Margolis and Myrskylä 2011. A Global Perspective on Happiness and Fertility. *Population and Development Review*. 37(1): 29-56.

Week 3: Linear Regression and Interaction Terms

Methods reading: Gordon textbook Chapters 7 and 8

Methods reading: Miller Chapter 9

Example reading: Joyner, Kara and J. Richard Udry. 2000. You Don't Bring Me Anything but Down: Adolescent Romance and Depression. *Journal of Health and Social Behavior* 41(4): 369-391.

Week 4: Regression Diagnostics, Logit Models I

Methods reading: Gordon textbook Chapters 9-11.

Methods reading: Peng, Lee and Ingersoll. 2002. An Introduction to Logistic Regression Analysis and Reporting. *The Journal of Educational Research*. 96(1): 1-14.

Example reading: Manning et al. 2000. The Relationship Context of Contraceptive Use at First Intercourse. *Family Planning Perspectives* 32(3):104-110. *Journal of Marriage and Family* 71: 174-186.

Week 5: Logit Models II

Methods reading: Pampel, Fred C. 2000. Logistic Regression: A Primer. Thousand Oaks, California. Sage Publications.

Example reading: Xie, Yu and Haiyan Zhu. 2009. Do Sons or Daughters Give More Money to Parents in Urban China? *Journal of Marriage and Family* 71: 174-186.

Week 6: Midterm.

No readings.

Week 7: Survival Analysis I

Methods reading: Clark et al. 2003. Survival Analysis Part I: Basic Concepts and First Analysis. *British Journal of Cancer* 89:232-238.

Example reading: Yoshikawa et al. 2007. Violent Recidivism among Mentally Disordered Offenders in Japan. *Criminal Behavior and Mental Health* 17:137-151.

Week 8: Survival Analysis II

Methods reading: Bradburn et al. 2003. Survival Analysis Part II: Multivariate Data Analysis- An Introduction to Concepts and Methods. *British Journal of Cancer* 89:431-436.

Methods reading: Allison. 1982. Discrete Time Methods for the Analysis of Event Histories. *Sociological Methodology* 13:61-98.

Example reading: Margolis, Rachel and Mikko Myrskylä. 2015. Parental Well-being Surrounding First Birth as a Determinant of Further Parity Progression. *Demography* 52(4):1147-1166.

* Note: Also see online appendix.

Week 9: Causality

Methods reading: Morgan, Steve. 2002. Should Sociologists Use Instrumental Variables?

Example reading: TBA

Week 10: Fixed Effects

Methods and Example reading: Geronimus and Korenman. 1992. The Socioeconomic Consequences of Teen Childbearing Reconsidered. *The Quarterly Journal of Economics* 107(4):1187-1214.

Week 11: Advanced Topics, Data Challenges

Example reading: Leventhal and Brooks-Gunn. 2003. Moving to Opportunity: An Experimental Study of Neighborhood Effects on Mental Health. *American Journal of Public Health*. 93(9):1576-1582.

Weeks 12-13: Workshop Student Projects

Important Information for Students

A Note on Plagiarism: Students must write their assignments in their own words. Whenever students take an idea from another author, they must acknowledge their debt both by using quotation marks where appropriate and by proper referencing such as footnotes or citations. Plagiarism is a major scholastic offence. See Scholastic Offences (below) for the link to Scholastic Discipline regulations.

Plagiarism Checking: All required papers may be subject to submission for textual similarity review to the commercial plagiarism detection software under license to the University for the detection of plagiarism. All papers submitted will be included as source documents in the reference database for the purpose of detecting plagiarism of papers subsequently submitted to the system. Use of the service is subject to the licensing agreement, currently between The University of Western Ontario and Turnitin.com: <http://www.turnitin.com>

Scholastic Offences: Scholastic offences are taken seriously and students are directed to read the appropriate policy, specifically, the definition of what constitutes a Scholastic Offence, at the following web site: http://www.uwo.ca/univsec/appeals_discipline/index.html

Accommodation

Only in exceptional circumstances may a student be provided special accommodation in the completion of a course requirement (i.e., exams, papers). To request a one-time accommodation (brief illness, family emergency), the student should inform the professor when they are able so accommodation can be made. To request other accommodation(s), the student must first meet with the Graduate Chair to discuss options. Medical documentation, where required, will be kept on file in the Sociology graduate program office.

Completion of Course Requirements

Course requirements must be completed by the end of the term in which the course is offered (Fall-December 31; Winter-April 30, Summer-August 31). Only in exceptional circumstances may a student take additional time to complete the course requirements. In such a case, the student must first meet with the Graduate Chair to request permission to carry the incomplete. Medical documentation, where required, will be kept on file in the Sociology graduate program office. More details regarding incompletes are outlined in the Graduate Handbook: http://www.sociology.uwo.ca/graduate_handbook/course_information.html

Mental Health: Students who are in emotional/mental distress should refer to Health and Wellness Western for a complete list of options to obtain help:

http://uwo.ca/health/mental_wellbeing/

Health and Wellness: As part of a successful graduate student experience at Western, we encourage students to make their health and wellness a priority. Students seeking help regarding mental health concerns are advised to speak to someone they feel comfortable confiding in, such as their faculty supervisor, their program director (graduate chair), or other relevant administrators in their unit. *The Wellness Education Centre (lower level UCC)* assists students in finding mental health and other related resources best suited to their needs: <http://se.uwo.ca/wec.html> Western's School of Graduate and Postdoctoral Studies' *Living Well* website provides tips for thriving at grad school and other helpful information:

http://grad.uwo.ca/current_students/living_well/index.html

Western provides several on-campus health-related services to help you achieve optimum health and engage in healthy living while pursuing your graduate degree. For example, to support

physical activity, all students, as part of their registration, receive membership in [Western's Campus Recreation Centre](http://www.westernmustangs.ca/index.aspx?path=ims#): <http://www.westernmustangs.ca/index.aspx?path=ims#> Numerous cultural events are offered throughout the year. Also, we encourage you to check out the Faculty of Music web page <http://www.music.uwo.ca/>, and our own McIntosh Gallery <http://www.mcintoshgallery.ca/>.

Data Analysis Project Instructions

The data analysis project is designed to help students develop a research question, conduct data analysis, make tables, and write and present the method and results. There are graded written and oral components. The idea is to use the skills gained throughout the semester.

Choosing a Topic, Data, and Method

- Choose a topic in your research area. It can be part of a thesis. It can be something you started for another course, or an idea you had a long time ago. It can be a paper that you are getting ready to send for publication.
- The project **cannot** be based on a paper that is co-authored. This must be YOUR work alone.
- Choose your data carefully. You will want data that will help you answer your question. You also want data that you have easy access to and that you can start working with right away. If you don't have a data set, you can browse publically available data sets ICPSR. <http://www.icpsr.umich.edu/icpsrweb/ICPSR/access/index.jsp>
- The project **must** use one of the methods that we are covering in the course. Choose a question that you can answer with one of these methods.

Evaluation

The data analysis project will comprise 55% of your final grade for the course. There are 3 graded components:

- 1) Research Questions and Data Source (10% of final grade).** Students will submit their research question and data source on **February 28**. This should be 1-2 pages. Explain what data you are using to answer your questions and why the data are well suited to the research question at hand.
- 2) Oral Presentation (20% of final grade).** Students will present their projects to the class on **April 4 and April 11**. Students should prepare 10-13 minute presentations on their research project, with 10-20 slides. This is a similar length presentation to those for most conferences and therefore is good practice. The presentation should highlight all the important parts of the paper: research questions, contributions, data, method, results, discussion. The whole audience should be able to see the text of your slides. Do not use text smaller than size 24.

Practice your presentation: A good presentation is one that you have practiced! Do it for your friends and have them tell you whether your slides and talk are clear.

Students will sign up for a presentation date by mid-semester. Presenting earlier in the course gives you time to incorporate the comments from your classmates into your project before submitting the final project on the last day of class.

- 3) Final Written Version of Project (25% of final grade).** On **April 23, 10am**, students will turn in the following sections of a research paper: Research Questions, Contributions, Data, Methods, Results, Discussion, Tables, and References.

Instructions for Final Project

The length of the final paper should be between 15 and 20 pages, with text double spaced. Including text, tables, figures, and references, the paper MUST NOT EXCEED 25 pages. No title page is necessary. Hand in a printed copy of the paper.

The paper should look professional and should be free of spelling, grammatical, and formatting errors. Use Times New Roman font (size 12) and double space the paper. Align the text on the left side of the page. All tables should go at the END of the paper, after the references.

The final paper should include the following sections and answer the following questions.

Research Questions

What research question/s do you answer in this paper?

Contributions

Explain succinctly what your project adds to the literature on your topic and why it's important.

Data

Data: What data do you use? Where did you get it? What are the strengths of the data for your research project? Any weaknesses should either be discussed in the data or with the limitations in the discussion.

Analytic sample: How many respondents are there in the data? How many in your analytic sample? If you are not using the whole sample, why did you drop respondents? How do you handle missing data?

Measures: What are the variables you chose to use? Justify including the variables you do and why you code them in the way you do.

Method

What is your analytic approach? For this project, you must use one of the methods that we cover in this course. Make sure to choose a research question that you can answer with one of these methods and that you do not need something more complex. Show the equation that you use for your multivariate model, including all variables that you use.

Results

Explain the results of your analysis. Start with the sample characteristics and explain why the descriptive results warrant the multivariate approach. Use the Miller text to help you decide how to talk about your results. You may also want to use an example paper to decide how much text to spend on your results. Make sure to tell the reader what tables or figures he/she should see for the results that you are discussing.

Discussion

What are the conclusions of your analysis? How do they contribute to what we know about this topic? Are they interesting? Important? How do they relate to other research on the topic? If they are different, discuss why they might differ.

Discuss the limitations of your project.

Tables and/or Figures

Your tables should show 1) That there is interesting variation in your dependent variable by your independent variable of interest, 2) Sample characteristics of your analytic sample, 3) Results from your multivariate analysis. All tables and figures should be formatted like those in the Miller text. All tables and figures should have clear titles explaining what is being shown.

References

Include a list of references, in the style used by ASA journals. Here are instructions. Refer to the *ASA Style Guide* (4th ed., 2010) or any issue of *American Sociological Review* or the *Journal of Health and Social Behavior* for additional examples.

References are presented in a separate section headed “REFERENCES.” All references cited in the text must be listed in the reference section, and vice versa. Publication information for each must be complete and correct. List the references in alphabetical order by authors’ last names; include first names and middle initials for all authors when available. List two or more entries by the same author(s) in order of the year of publication. When the cited material is not yet published but has been accepted for publication, use “Forthcoming” in place of the date and give the journal name or publishing house. For dissertations and unpublished papers, cite the date and place the paper was presented and/or where it is available. If no date is available, use “n.d.” in place of the date. If two or more cited works are by the same author(s) within the same year, list them in alphabetical order by title and distinguish them by adding the letters a, b, c, etc., to the year (or to “Forthcoming”). For works with more than one author, only the name of the first author is inverted (e.g., “Jones, Arthur B., Colin D. Smith, and James Petersen”). List all authors; using “et al.” in the reference list is not acceptable. References for data sets should include a persistent identifier, such as a Digital Object Identifier (DOI). Persistent identifiers ensure future access to unique published digital objects, such as a text or data set. Persistent identifiers are assigned to data sets by digital archives, such as institutional repositories and partners in the Data Preservation Alliance for the Social Sciences (Data-PASS). Refer to the *ASA Style Guide* (4th ed., 2010) for additional examples:

1. *Books:*

Bernard, Claude. [1865] 1957. *An Introduction to the Study of Experimental Medicine*. Translated by H. C. Greene. New York: Dover.

1. Mason, Karen O. 1974. *Women’s Labor Force Participation and Fertility*. Research Triangle Park, NC: National Institutes of Health.
2. U.S. Bureau of the Census. 1960. *Characteristics of Population*. Vol. 1. Washington, DC: U.S. Government Printing Office.

2. *Periodicals:*

Goodman, Leo A. 1947a. “The Analysis of Systems of Qualitative Variables When Some of the Variables Are Unobservable. Part I—A Modified Latent Structure Approach.” *American Journal of Sociology* 79:1179–1259.

1. Goodman, Leo A. 1947b. “Exploratory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models.” *Biometrika* 61:215–31.
2. Szelényi, Szonja and Jacqueline Olvera. Forthcoming. “The Declining Significance of Class: Does Gender Complicate the Story?” *Theory and Society*.

3. *Collections:*

Sampson, Robert J. 1992. “Family Management and Child Development: Insights from Social Disorganization Theory.” Pp. 63–93 in *Advances in Criminology Theory*. Vol. 3, *Facts, Frameworks, and Forecasts*, edited by J. McCord. New Brunswick, NJ: Transaction.

4. *Dissertations:*

Charles, Maria. 1990. “Occupational Sex Segregation: A Log-Linear Analysis of Patterns in 25 Industrial Countries.” PhD dissertation, Department of Sociology, Stanford University, Stanford, CA.

5. *Web*

sites:

American Sociological Association. 1997. “Call for Help: Social Science Knowledge on Race, Racism, and Race Relations” (ASA Action Alert, October 15). Washington, DC: American Sociological Association. Retrieved October 15, 1997 (<http://www.asanet.org/racecall.htm>).

6. Kao, Grace and Jennifer Thompson. 2003. “Racial and Ethnic Stratification in Educational Achievement and Attainment.” *Annual Review of Sociology* 29:417–42. Retrieved October 20, 2003 (<http://arjournals.annualreviews.org/doi/abs/10.1146/annurev.soc.29.010202.100019>).

7. *DataSets:*

Deschenes, Elizabeth Piper, Susan Turner, and Joan Petersilia. *Intensive Community Supervision in Minnesota, 1990–1992: A Dual Experiment in Prison Diversion and Enhanced Supervised Release* [Computer file]. ICPSR06849-v1. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2000. doi:10.3886/ICPSR06849.