



SW 862: Categorical Data Analysis
Fall 2020

The First Book

Open it.

Go ahead, it won't bite.
Well ... maybe a little.

More a nip, like. A tingle.
It's pleasurable, really.

You see, it keeps on opening.
You may fall in.

Sure, it's hard to get started;
remember learning to use

knife and fork? Dig in:
you'll never reach bottom.

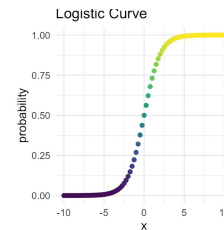
It's not like it's the end of the world—
just the world as you think

you know it.

— Rita Dove

"Mathematics is the art of giving the same
name to different things."

— Henri Poincare



Why can't numbers be beautiful too?

We all talk of beautiful words, art, buildings
and they're not part of the natural world, either.
An x in Algebra is no more abstract than
an idea in philosophy, just more useful.

... Of course, I know there is no great beauty in
a single number, in a *four* or a *seven* or an
eight, but it is the same with the alphabet.
Where is the wonder in a *b* or a *k* or a *t*? It is
only the combinations, the meanings, the
relationships between the letters that make the
words and sounds we love.

— And so, why can't my numbers be beautiful
to me? Why the scorn, the doubt in your face?
Do you think I am brittle and dusty as old
paper?
Look again. See the numbers shine in my
eyes.

— Eveline Pye

"The best thing about being a statistician is
that you get to play in everyone's backyard."

— John Tukey

COVID-19 Statement

For the safety of all students, faculty, and staff on campus, it is important for each of us to be mindful of safety measures that have been put in place for our protection. By returning to campus, you have acknowledged your responsibility for protecting the collective health of our community. Your participation in this course on an in-person/hybrid basis is conditional upon your adherence to all safety measures mandated by the state of Michigan and the University, including maintaining physical distancing of six feet from others, and properly wearing a face covering in class. Other applicable safety measures may be described in the [Wolverine Culture of Care](#) and the [University's Face Covering Policy for COVID-19](#). Your ability to participate in this course in-person/hybrid may be impacted by failure to comply with campus safety measures. Individuals seeking to request an accommodation related to the face covering requirement under the Americans with Disabilities Act should contact the [Office for Institutional Equity](#). If you are unable or unwilling to adhere to these safety measures while in a face-to-face class setting, you will be required to participate on a remote basis. I also encourage you to review the [Statement of Student Rights and Responsibilities](#) and the [COVID-related Addendum to the Statement of Student Rights and Responsibilities](#).

Health-Related Class Absences

Please evaluate your own health status regularly and refrain from attending class and coming to campus if you are ill. You are encouraged to seek appropriate medical attention for treatment. School of Social Work students who miss class due to illness of any kind will be given opportunities to access course materials online or provided with alternative learning opportunities. Please notify me by email about your absence as soon as practical, so that I can make accommodations. Please note that documentation (a Doctor's note) for medical excuses is *not* required.

Recording Class

Audio and video recording of in-class lectures and discussions is prohibited without the advance written permission of the instructor. Students with an approved accommodation from the Office of Services for Students with Disabilities permitting the recording of class meetings must present documentation to the instructor in advance of any recording being done. The instructor reserves the right to disallow recording for a portion of any class time where privacy is a special concern. If the instructor chooses to record a class, they will decide which classes, if any, are recorded, what portion of each class is recorded, and whether a recording is made available on the course management website. On days when classes are recorded, students will be notified in advance that a recording will occur and be provided with an option to opt-out. Class recordings and course materials may not be reproduced, sold, published or distributed to others, in whole or in part, without the written consent of the instructor.

Credit Hours: 3
Prerequisites: None
Instructor: Andy Grogan-Kaylor, MA, MSSW, PhD, Profesor, SSW
Pronouns: he, his, him
Office: 3846 School of Social Work
Phone: 615-3369
Email: agrogan@umich.edu (best way to get in touch with me)
(please put SW862 in the subject line)
Office Hours: Mondays, 1-3, and by appointment.

Course Web site at <http://canvas.umich.edu>

Course Description and Content

Researchers are most commonly aware of methods that are suitable for continuous dependent variables (e.g. mental health scores), such as the use of ordinary least squares regression. However, many outcomes of interest to social workers, and other social researchers, are decidedly not continuous, but are dichotomous or binary in nature: entered the program versus did not enter the program; left the program versus stayed in the program; received a particular diagnosis; did not receive a diagnosis. Many researchers are familiar with the basics of logistic regression, yet do not have a grounding in some of the intricacies of logistic regression, such as generating predicted probabilities, or using interaction terms in a categorical model, which can lead to clearer and more accurate reporting of results. Further, the basic logistic regression model serves as the foundation for a wide variety of more advanced statistical approaches that can help advance social work research. Study of the logistic regression model can lead to variations of logistic regression such as logistic regression for ordered variables, or multinomial logistic regression where there are more than two categories of the outcome variable (e.g. multiple forms of family violence). An understanding of logistic regression also helps to motivate understanding of models for count data such as the Poisson and negative binomial model suitable for studying counts of events such as incidence of disease or incidence of violence. Lastly, categorical data models serve as the foundation for event history models that are used to study the timing of events, such as the timing of program entry, program departure, or receipt of a diagnosis.

The treatment of these topics involves some discussion of the underlying statistical theory, but is much more focused on the application of these models with real data. Stata is intended to be the primary statistical software used in this course, though it is likely that some occasional reference will be made to R syntax and ways of doing things. The statistical models studied in this course are understood to transcend any one particular statistical software program.

Method of Instruction

Instruction is conceived as a mixture of in class lecture and discussion and time spent in the classroom working through actual data problems with statistical software in order to convey relevant content.

Required Readings

My presentation of the material for the course will draw on the Stata documentation and on Regression Models for Categorical Dependent Variables, 3rd Edition by Long and Freese, which you should purchase from Stata Press.

All other readings for this course will be made available using <http://canvas.umich.edu>.

Course Requirements

- 1) **Three 8-10 page “mini-papers” focusing upon the statistical analysis on a research question of interest using secondary or other available data.** The first of these papers is intended to be a logistic regression analysis and the second paper is to be a paper using an extension of logistic regression such as an ordered, multinomial, or count model. A third paper should employ an event history model. **Note that each of these models makes demands of the data: a logistic regression model requires a binary outcome; other models require ordered, multinomial or count outcomes; and event history models require data with information on the timing of events.** At a minimum, each paper should include some basic demographic controls in addition to one or two independent variables of theoretical or substantive interest. At times, your demographic variables may in fact be the relevant independent variables of theoretical interest. The papers will be due at different points over the term. Consultation with the professor and other students will be available. Quality of writing, and clarity of presentation, will be one of the criteria used in grading.¹ Please make sure that you cite other’s work properly and avoid plagiarism. Plagiarism—when discovered—will be dealt with severely. Each paper should include: a review of at least 5 key pieces of background literature²; descriptive statistics; a multivariate analysis³; a well-documented and nicely formatted and labeled graph illustrating some bivariate aspect of your data; and a discussion of your findings and their implications. Further details of these assignments will be forthcoming and will also appear on the class Web site.
- 2) **Class participation.** You will learn more the more you participate. Because this is a doctoral course it is important to participate in some way, and especially to ask questions. I recognize that there are many variations that affect the rate of participation in this course, but a portion of your grade will depend on class participation. **[20% of grade]**
- 3) **Attendance.** You are expected to attend each class and lab session. If you miss a class session for any reason, please see me to make sure that you are grasping the material. In particular, you will be expected to complete all lab assignments, unless we make other arrangements.

¹ Please note that for purposes of this course, plagiarism consists of six or more consecutive words, taken from another source without proper attribution. Failure upon my part to detect plagiarism does not imply approval of plagiarism.

² Please note that this literature review is not intended to be exhaustive, merely to pave the way for the essential argument of the paper.

³ While I have no set format for presentation of tables of results, please pay some attention to formatting your results, and do not simply paste unformatted statistical output directly into your paper.

- 4) **Lab Work.** We will spend some time every week doing lab work⁴. The purpose of the lab is to give you hands-on experience with the building blocks of categorical data analysis as well as working with the specific kinds of data structures pertaining to this type of analysis. Most of the learning in lab occurs in the doing of lab. Lab is intended to be a collaborative endeavor where students work with the instructor and with each other. I will not directly grade your lab work, or be able to provide written feedback on your lab work. I am always happy to discuss lab work during lab, during office hours, or over e-mail.

A note on work handed in late: Most students turn in work in accordance with class deadlines. In order to be fair to the majority of students, I have developed the following policy: late work will be graded down by half a grade a day unless prior arrangements for an extension have been made with me. I very much understand that extenuating circumstances may arise which make it difficult to turn in work on time. All I am asking you to do is to communicate with me if you need some kind of extension so that we can work out an arrangement that is mutually agreeable.

Grades

Class Participation	20%
Choosing First Data Set	7%
First Mini-Paper	20%
Choosing and Preparing Second Data Set	7%
Second Mini-Paper	20%
Choosing and Preparing Third Data Set	6%
Final Mini-Paper	20%

Grading Scale⁵

"A" grades are given for exceptional individual performance and mastery of the material. The use of "A+", "A", and "A-" distinguishes the degree of superiority. "B" grades are given to students who demonstrate mastery of the material, at the level expected for the course. "B+" is used for students who perform just above the mastery level but not in an exceptional manner. "B-" is used for students just below the mastery level. "C" grades are given when mastery of the material is minimal. A "C-" is the lowest grade which carries credit. "D" grades indicate deficiency and carry no credit. "E" grades indicate failure and carry no credit.

Grading Criteria for Written Assignments⁶

The three mini-papers will be graded using the following criteria:

- 1) The quality of the analysis and depth of understanding of the concepts, ideas, and information presented.
- 2) The clarity of expression and organization of the paper – i.e., a logical order to the presentation of its major themes.

⁴ You may use your own laptop to connect to <http://virtualsites.umich.edu>. NB that while virtualesites has considerably improved over the years, it can be a “buggy” experience. Let me know if you need help with this.

⁵ Adapted from the MSW student guide

⁶ Adapted from a statement by Michael Reisch.

- 3) The appropriate use of references and resources, and the variety/range of resources referenced.
- 4) The use of proper grammar and the overall professional presentation of the paper.

Note: Please take time to proofread your writing and make sure you submit a second or third draft.

Data Sets

All of the written paper projects for this course require you to make use of previously collected data, usually publicly available secondary data, for your analyses. I will work with you to identify suitable data.

It may be the case that you have access to data that is not publicly available, for example data from a research project, or agency, with which you are involved. It is perfectly acceptable to use such data, but you must have a copy of this data in hand at the time that you turn in your one paragraph on each study.

Software

- Stata (available virtualsites, and from <https://www.stata.com/>) (in class exercises will focus on the use of Stata although the statistical concepts covered transcend any one statistical software package)

Week	Date	Topic	Reading(s)	Assignments	Data
1	8/31/2020	Welcome & Introduction			
2	9/7/2020	Labor Day Break			
3	9/14/2020	Review of OLS / Introduction to Stata and Stata graphing	Long & Freese, Chapter 2	one paragraph description of the data and question you will work on for the first paper	Simulated Client Data
4	9/21/2020	Contingency Tables	Stata help <code>tabulate</code> focusing on 2 way tables.		French Skiers
5	9/28/2020	Linear Probability Model, Probit, Logistic Regression (GLM)	Long & Freese, Chapter 4 (Skim) then Chapter 5.1-5.4		General Social Survey
6	10/5/2020	Linear Probability Model, Probit, Logistic Regression (GLM) (2)	Same		General Social Survey
7	10/12/2020	Lab Day 1			
8	10/19/2020	Ordered and Multinomial Models	Long & Freese Chapter 7 & 8	first mini-paper (logistic)	General Social Survey
9	10/26/2020	Count Models	Long & Freese Chapter 9	one paragraph description of the data and question you will work on for the second paper	General Social Survey, possibly National Survey of Children's Health
10	11/2/2020	Multilevel Perspectives on CDA	TBD		General Social Survey
11	11/9/2020	Lab Day 2			
12	11/16/2020	Event History and the	Stata documentation	second mini-paper	Stata pharmaceutical drug testing data

		Timing of Events	on stset and stcox	(extension of logistic)	
13	11/23/2020	THANKS-GIVING BREAK		one paragraph description of the data and question you will work on for the third paper	
14	11/30/2020	Bayesian Perspectives on Categorical Data Analysis	Stata help bayes https://agrogan.shinyapps.io/Thinking-Through-Bayes/		General Social Survey
15	12/7/2019	Lab Day 3 (last day of classes)			
		week after classes end		final mini-paper (event history)	