

Course Syllabus for Sociology 63993
Graduate Statistics II
Spring 2009

Instructor Richard Williams
741 Flanner (574-631-6668)
Office Hours: Thursday 10:30-12:00 and by appointment
Email: Richard.A.Williams.5@ND.EDU
Personal Web Page: <http://www.nd.edu/~rwilliam/>
Course Web Page: <http://www.nd.edu/~rwilliam/xsoc63993/index.html>

TA: Yuting Wang
839 Flanner (574-383-2589)
Office Hours: Tuesday 10:30-12:00 and by appointment
Email: ywang7@nd.edu

Time and Place Class MW 11:45-1:00, 242 DeBartolo
Lab F 1:30-2:45, 228 DeBartolo

Readings and Texts

Multiple regression in practice, by Berry and Feldman (1985) [Required]

Multiple Regression: A Primer, Paul Allison (1999) [Required]

Online Readings Packet for Sociology 6393, compiled by Williams [available in electronic form; most are recommended but not required]

Statistics with Stata, updated for Version 10, by Lawrence Hamilton (2009) [suggested, not required]

Spss 16.0 Statistical Procedure Companion, by Marija J. Norusis (2007) [suggested, not required]

The required books are in the bookstore. The readings packet includes articles and book chapters from several sources. Additional suggested readings may be placed on reserve or made available on the World Wide Web.

MOST COURSE NOTES, HANDOUTS, HOMEWORK ASSIGNMENTS, AND ANSWER KEYS WILL BE AVAILABLE ON THE WORLD WIDE WEB. Unless I say otherwise, you are expected to print out notes beforehand and bring them with you to class. Sometimes, we may be covering topics a little ahead of schedule, so you should also try to print things out at least a week ahead. The course materials are available at:

<http://www.nd.edu/~rwilliam/xsoc63993/index.html>

If you are having trouble accessing or printing the handouts, you should speak to me or the TA immediately. Don't wait until the last minute to print something.

Classroom Format/Readings/Exams, Homework, Paper

Readings & Lecture. I will lecture heavily from the handouts I give you, deviating as the need arises. You should at least skim through handouts before class and be prepared to ask questions. Occasionally I will give you short exercises to work on in class.

Obviously, the readings can go into much more depth than I can in class. They also provide additional examples. You may find that the course notes are adequate for most of your needs, but at a minimum I would suggest going over the excellent and concise Allison and Berry and Feldman books carefully. You should definitely do the recommended readings whenever you want additional help. I will indicate when I think the readings will be especially useful. Parts of the readings sometimes go well beyond what is covered in this class, so do not be frustrated if you can't always follow them.

SPSS and Stata. In the past, I primarily used SPSS in this course. However, more and more faculty and students are using both SPSS and Stata, and some use Stata exclusively. Both have their strengths and weaknesses. I think it is good to be familiar with both packages, so both will be used this semester. The suggested Stata and SPSS books are very useful for understanding those programs but they also serve as good, basic, non-mathematical statistics texts.

Examinations & Homework. There will be three open-book examinations, each counting for 20% of your grade and 60% altogether. Exams have traditionally been on Fridays from 1:00-3:00 but we can change that if necessary. Occasionally I may start on new material before I give the exam (e.g. start exam 3 material before we have had exam 2) in order to give you adequate time to prepare for the exam. Things like the ever-shifting midterm and Easter breaks sometimes make it hard to schedule exams and review sessions at what I would consider the optimal time.

There will also be 10 homework assignments. These will count for 10% of your grade. In effect, each assignment is treated as receiving either an "A" or an "F". You do not have to get all parts of an assignment right, but you do have to make a good faith effort to complete it. You should give the homework to the TA (not me) by 5:00 on the due date. It is very important to not fall behind on the material so assignments are required to be handed in on time.

Paper. There is also a term paper, worth 30% of your grade. The paper should be a short (10-15 pages) write-up of data analysis on the sociological problem of your choice. The goal of the paper is to demonstrate your ability to formulate a research hypothesis, and appropriately apply the correct statistical technique to test that hypothesis. In addition, you will be expected to discuss any data handling or estimation problems you encounter in your analysis (such as missing data, multicollinearity, non-linear relationships, etc.). I STRONGLY encourage you to discuss your papers with me or the teaching assistant as you are working on them. Toward that end, **I want to see a rough draft (or at least a detailed outline) of your paper before the midterm break.** [NOTE: Please keep the drafts/outlines concise though – I'd prefer not to read an entire 30-50 page paper that you wrote for another class and are now thinking about redoing with improved methods! 10 pages or less is generally adequate.] You are also encouraged to meet with me or the TA as you work towards your final version. All papers (draft and final) should be submitted in both printed and electronic version to both myself and the TA. If all professors

involved agree, you are welcome and even encouraged to submit this paper to more than one class; expectations will of course be higher in such cases.

You may use any data set that you like. One source that many students have found helpful in the past is the General Social Survey. For information on the GSS, see

<http://www.norc.org/GSS+Website/>

Links to several other data sources, as well as information on available support services, can be found at Notre Dame's Social Research Data Services web page:

<http://www.nd.edu/~datamgt/>

The paper is due on the last regular class day, but you are welcome and encouraged to hand it in sooner.

Lab: During the lab sessions, the teaching assistant will help you with any problems you are having with the homework or the course. S/he will also spend time discussing data sets that are available and how to access and use them. Feel free to offer suggestions to the TA on the possible content of lab sessions.

Especially if we are falling a little behind, I may occasionally take over the lab sessions. I may also schedule additional optional review sessions near exam time.

Schedule: I have tried to divide the course up into three logical components. Whether I have accurately estimated the time required for each component (especially in the final third of the course) remains to be seen. Hence, the attached schedule is hopefully more or less firm, but I reserve the right to make modifications (reorder the presentation of topics, move exams up or back a little). Also, the due dates for homeworks may be adjusted depending on how we are progressing in class.

Term paper suggestions

CHOOSE A TOPIC THAT IS SUBSTANTIVELY INTERESTING TO YOU. Make sure that

- The topic is subject to empirical investigation
- Data are readily available. You can use the General Social Survey if you want. You are also free to use any other convenient data set. Do not pick a data set that will take months to obtain or get ready. Sometimes people pick a sub-optimal but convenient data set for this paper and then re-do the analysis when the data they want are available.
- The topic is not too ambitious. This is a short paper, not a dissertation or even a Master's thesis. On the other hand, this could well be a start for your Master's thesis or a journal article.

SPEND SOME TIME MOTIVATING INTEREST IN YOUR TOPIC.

- Review the relevant literature, at least briefly. People sometimes do not take this part seriously enough. You don't want to spend a lot of time researching something that we already know the answer to.
- Explain why this is an interesting and important topic, and why your ideas are reasonable.
- Specify models and/or hypotheses. Make sure that these are stated in such a way that they can be empirically tested. It must be possible for the hypotheses or model to be dis-confirmed.
- Make clear what your contribution is. Sometimes people have great literature reviews, but you are left wondering why anyone would want to do additional research on the topic. You could identify gaps and weaknesses in past research. It is also often helpful to discuss competing theories, and explain how you will adjudicate between them. In general, don't make it sound like we already know the answers and the results are a foregone conclusion.

APPLY THE LESSONS YOU HAVE LEARNED FROM THIS COURSE. It might help to review the homework assignments; the same sorts of questions that were asked in them could be asked about your specific problem. For example:

- Is there missing data? If so, what is the nature of the missing data? How should missing data be handled?
- Do heteroscedasticity or multicollinearity appear to be problems? If so, how will you deal with them?
- How will you decide what variables should be in the equations? Theories may lead to conflicting model specifications. Test whether a variable, or a set of variables, belong in the equation.
- Are there subgroup differences, e.g. is the model different for men than for women? If so, how will you deal with these differences (e.g. will you use interaction effects)?
- Are there non-linear effects? If so, include them in the model.
- Will you specify a complete path model? If so, explain the logic behind the model, its substantive implications, and how you will estimate it.
- Is OLS regression appropriate for this problem? If not, explain why, and use a more appropriate technique (e.g. logistic regression)

FINALLY, DISCUSS THE SUBSTANTIVE IMPLICATIONS OF YOUR FINDINGS.

Sociology 63993 — Detailed Outline

[Approximate numbers of days to be spent on each topic are in brackets]

I. Multiple regression in practice. Review of regression; problems with data; common violations of OLS assumptions.

- Overview [1]
- Review of multiple regression [1]
- Multicollinearity [1]
- Missing data [1]
- Measurement error in variables (random and nonrandom; reliability); Scale construction [1]
- Outliers; Robust regression techniques [1]
- Serial Correlation and Heteroscedasticity [1]

II. Causal modeling. Model specification and misspecification; choosing between models; Specifying the correct form of relationships; The logic of causal order; path analysis; use of standardized versus metric coefficients in causal modeling

- The logic of causal order [2]
- Specification error: Omitted and extraneous variables [1]
- Comparing and testing nested models; Comparing and testing models across populations; Interaction effects [3]
- Modeling nonlinear relationships [1]
- Intro to path modeling [1]

III. Advanced causal modeling and statistical techniques. More on path analysis; Non-recursive models; Techniques to use when OLS regression is not appropriate

- Advanced path analysis. Includes sidelights on the use of standardized versus metric coefficients in causal modeling; R^2 ; Recursive path models [1-2]
- Logistic regression, Multinomial Logit, Ordinal regression [2-3]
- Non-recursive models. Instrumental variables, indirect least squares, 2 stage least squares [1]
- Brief overview of Other Advanced techniques – Manova, LISREL, Event History Analysis and Hierarchical Linear Modeling [2-3]

Tentative Schedule
Sociology 63993, Graduate Statistics II
Spring 2009

<i>Date</i>	<i>Topic</i>	<i>Assignment</i>	<i>Readings</i>
1/14	Overview		PA, ch. 1-6 (see especially ch. 1 & 3)
1/19	Review of multiple regression	H#1 due 1/28	BF, ch. 1; SPSS ch. 13, Stata ch. 6 RP, "Introduction"
1/21	Multicollinearity		BF, ch. 4, PA, ch. 7
1/26	Missing data	H#2 due 2/4	RP, "Missing Data"
1/28	Measurement error; Scale construction		BF, ch. 3; SPSS chs.17, 18; Stata ch. 12 RP, "Measurement error"
2/2	Outliers; Robust Regression		BF ch. 6; Stata ch. 7, 9; RP, "Outliers"
2/4	Heteroskedasticity, Serial Correlation	H#3 due 2/11	RP, "Heteroskedasticity"
2/9 and/or 2/11	Review, Catch up; Additional topics may be brought up if felt necessary or if time permits		
2/13	Exam I		
2/16	The logic of causal order		RP, "Logic of Causal Modeling"
2/18	The logic of causal order	H #4 due 2/25	
2/23	Specification error; Testing models & constraints		BF, ch. 2
2/25	Group comparisons	H#5 due 3/4	
3/2	Modeling & interpreting interaction effects		
3/4	Modeling & interpreting interaction effects	H#6 due 3/18	RP, "Interaction effects"
3/4	FIRST DRAFTS/OUTLINES DUE		
MIDTERM BREAK MARCH 7 – MARCH 15			
3/16	Modeling Nonlinear relationships		BF, ch. 5; PA, ch. 8; Stata ch. 8 RP, "Nonlinear Relationships";
3/18	Intro to path modeling	H # 7 due 3/25	RP, "Intro to Path Analysis"
3/23 and/or 3/25	Review, Catch up; Additional topics may be brought up if felt necessary or if time permits		
3/27	Exam II		
3/30	Structural coefficients/Evils of standardization/ R Square	H#8 due 4/8	
4/1	Logistic regression		SPSS chs. 15, 20; Stata ch. 10
4/6	Logistic regression	H # 9 due 4/15	RP, "Logistic Regression & Other Alternative Regression Models"
4/8	Multinomial & Ordered Logit Models		
EASTER BREAK APRIL 10 – APRIL 13			
4/15	Nonrecursive models		RP, "Nonrecursive models"
4/20	Overview of MANOVA, LISREL		SPSS ch. 23; RP, "Brief Overview of Other Advanced Methods"
4/22	Other advanced methods		
4/27, 4/29	Review, Catch up; Additional topics may be brought up if felt necessary or if time permits	H#10, Papers due 4/29	
5/6	Final Exam, 8:00-10:00 (Subject to Change)		

BF = Berry and Feldman (1985) Multiple regression in practice
PA = Paul Allison (1999) Multiple Regression: A Primer
RP = Online Readings Packet for Sociology 63993

Stata = Lawrence Hamilton (2009) Statistics with Stata, Updated for Version 10
SPSS = Marija J. Norusis, SPSS 16.0 Statistical Procedures Companion