

SYA 5407: Advanced Quantitative Techniques* – Fall 2007

* A.K.A. “Generalized Linear Models” or “Models for Limited Dependent Variables”
8 – 9:15am; Monday & Wednesday; 219 New Classroom Building (next to Bellamy)

Instructor: Dr. John Reynolds, 223 Pepper Center Bldg, 644-4321, john.reynolds@fsu.edu

Office Hours: Tuesdays and Wednesdays, 10-11:30a; *call first to verify I'm not in another meeting*

Course TA: Nicole Taylor, 506 Bellamy Bldg, 644-8161, cnt05d@fsu.edu

Web page: Blackboard site on campus.fsu.edu

Course Description

This course reviews four major regression models for limited dependent variables (RMLDVs), i.e., models for categorical outcomes, ordered outcomes, count outcomes, and event outcomes. Upon successful completion of the course you will be able to: (a) read and evaluate published or presented research that uses RMLDVs; (b) use Stata to reproduce existing RMLDV results from the textbook and create your own results from secondary data; and (c) translate the results into more useful summaries through tables and figures of predicted outcomes. You will be assessed with five assignments, five in-class quizzes, 15 online surveys, and a take home final exam. Completion of SYA 5406 is required for admission to the course; other graduate-level multiple regression courses may fulfill this prerequisite with instructor's approval. The course also requires skills that would be covered in a basic college algebra course.

Course Learning Objectives

1. Read and understand published research using RMLDVs.
2. In a given situation, use appropriate criteria to determine if RMLDV is appropriate, and if so which RMLDV should be used.
3. Use Stata syntax files to generate regression results for each type of RMLDV.
4. Interpret the output from Stata, both in terms of individual coefficients and in terms of measures of a model's goodness of fit.
5. Translate Stata's output into more useful formats, e.g., tables/graphs of probabilities.
6. Demonstrate good research skills through research logs.

Required Texts and Materials

Hoffman, John P. 2004. *Generalized Linear Models: An Applied Approach*. New York: A. B. Longman. ISBN: 0205377939.

Dunteman, George H. and Moon-Ho R. Ho. 2006. *An Introduction to Generalized Linear Models*. Thousand Oaks, CA: Sage. ISBN: 0761920846

Additional readings: Attached list of articles to be downloaded from Blackboard.

Optional, but useful: Scientific calculator, such as Texas Instruments' TI30X or TI34X (~\$20).

Recommended Text

J. Scott Long and Jeremy Freese. 2003. *Regression Models for Categorical Dependent Variables Using Stata*. College Station, TX: Stata Press. ISBN: 1881228827.

Course Requirements

1. Five assignments will be required throughout the semester. The assignments include a computer-based analysis, your interpretation of the results, Stata syntax and Stata output. Each assignment is worth 12 points, and together they account for 60 percent of your course grade. The assignments are due by midnight of the due date and should be submitted electronically through the course Blackboard site. It is your responsibility to confirm that you have successfully submitted the assignment. Late assignments will be penalized 1 point per day they are late (see grading rubric on next page), unless you have written beforehand with a valid excuse for the delay and a specific plan for when it will be submitted.
2. There will be five short in-class quizzes that assess your knowledge of the material and/or your ability to read published research that uses RMLDVs. The five quizzes are each worth 3 points, and as a group they make up 15 percent of your grade. The grading criteria are described on the next page.
3. For each unit of the course (not each week), you will be responsible for reading two to three published research articles that demonstrate the statistical techniques relevant to the unit. I recommend you take notes that include the following: (a) a complete reference to the article; (b) a 1-2 sentence summary of the research topic and question; (c) the data set used in this paper, mentioning something about the population (e.g., adolescents born between 1958 and 1965); (d) the statistical modeling technique used in the paper that corresponds to the technique we cover in class; (e) the authors' rationale for using this technique; and (f) an example of an interpretation of the results from the relevant statistical analyses. I may draw from these readings in lecture, in the surveys, in quizzes, and in the final exam.
4. You will complete 15 online "surveys" throughout the semester on the course Blackboard site. The surveys serve a variety of functions and are graded on a credit/no credit basis. The only time you will not receive credit for these surveys is if you do not complete them or it is obvious you put no thought into your responses. Sometimes these surveys are to determine if you have the skills needed to proceed in the course; other surveys are to generate feedback for use in lectures; other surveys may follow a more traditional skills quiz format. To receive credit, complete the survey by midnight of the day before class on Wednesday (or ASAP for the 1st week of class). These are worth 1 point each and together make up 15 percent of your course grade.
5. The course will have a take-home cumulative final exam. The final exam is due by midnight of Wednesday, December 12. The final is worth 10 points.

Grading Summary

(5) Assignments @ 12	60 points	A	90 – 100% of total points
(5) In-class quizzes @ 3	15 points	B	80 – 89%
(15) Online surveys @ 1	15 points	C	70 – 79%
(1) Final Exam @ 10	10 points	D	60 – 69%
<u>TOTAL</u>	<u>100 points</u>	F	below

I also assign pluses and minuses according to standard FSU guidelines. For grades A through D, the bottom 3 points in each grade range will receive a "-" (e.g., 80 to 82 = B-). For a grade of B, C, or D, the top 3 points in each grade range will receive a "+" (e.g., 77 to 79 = C+).

Grading Rubrics

Assignments

- 12 pts – no errors or, at most, very minor mistakes in math or in wording; great job
- 11 pts – no errors but did not fully address what one of the questions was asking; or a single error in terms of interpretation or model/test execution; a good job overall
- 10 pts – two major errors in execution or interpretation, or incomplete answers; shows a good level of effort but inadequate attention to or knowledge of some details
- 9 pts – more than two major errors in execution or interpretation; suggests inadequate effort or serious confusion about the methods featured in the assignment
- 0 pts – if you copy another student's syntax or report, you will receive no credit
(1 pt deduction for late assignments per day they are late without prior authorization.)

Quizzes

- 3 pts – no errors or, at most, very minor mistakes in math or in wording
- 2 pts – one or two questions missed; you get the general idea, but not all the details
- 1 pt – half or more of the questions are missed; indicates large deficit in understanding
- 0 pts – the only way to get no credit for a quiz is in the case of an unexcused absence

Surveys

Surveys are graded on credit/no-credit basis, with credit given for on-time completion.

Final Exam

There are 10 questions on the final exam. Each question is worth 1 point.

Computer Assignments

Data Sets

The core of this course entails completion of five computer assignments: the first assignment is intended to get you up to speed using Stata to generate statistical output and to use Long and Freese's customized post-estimation commands for RMLDVs. The other assignments cover four major types of RMLDVs—logit/probit, ordered probit, Poisson/NBR, and event history analysis.

You'll submit three separate documents for each assignment: a Stata syntax file used to generate the output; the output; and an answer sheet including answers to the assignment's questions, any relevant tables/figures, and a summary of the data and measures you used in the assignment.

Throughout the class, I encourage you to work with your classmates when grappling with the text, when debugging syntax problems, or when attempting to reproduce the examples from lecture or either of our two textbooks. However, the work you complete on the assignment must be your own work, and each of you is responsible for analyzing a unique set of variables. Ideally the data set will be one with which you are currently working. I also have a large number of data sets handy for your use (NELS, NLSY, ECLS, NSCW, Census, etc.). The main requirement is that you analyze unique variables and that you estimate regression models that have significant results and require interpretation. Failure to analyze a unique topic or estimate models with significant findings will result in having to redo the assignment and/or receiving a lower grade (significant finding means $b \neq 0, p < .05$). *Note: If you do plan to use your own data set, it needs to be one with which I am familiar in order to make it possible for me to grade your work fairly.*

Where to Use/Purchase Stata: The sociology graduate student computer lab has multiple copies of Stata (and its companion product, StatTransfer). Unfortunately, the computer lab is not large enough to accommodate students from other programs. If you are not a sociology graduate student, you need to buy your own copy (I recommend purchasing StatTransfer, too). In any case, if you would like to obtain a copy for your personal computer or laptop, you can purchase either a one-year license or a perpetual license for Intercooled Stata v.10 as a graduate student for \$95 or \$155, respectively, at <http://www.stata.com/order/new/edu/gradplan.html>. Stata/SE is only available with a perpetual license and costs \$335. Either version will work for this class. The main limitation of Intercooled Stata is this: "Intercooled Stata allows datasets with up to 2,047 variables. The limit of observations is based on the amount of RAM in your computer. Intercooled Stata allows string variables to contain a maximum of 80 characters." For further information, consult <http://www.stata.com/order/options-e.html#difference>.

Problems with Stata: When you run into errors while running Stata, try these sources of help.

1. Consult the in-program Help menu and use the search function to examine the syntax rules for the type of command you are trying to run.
2. Ask a classmate for assistance with your error, many students have some experience using Stata and will be able to provide some assistance.
3. Ask your TA or me. If you want to ask Nicole or me about an error, you must bring both your syntax file and the output file showing the error (if asking for help via email, send both as attachments). Without these files we will not be able to provide you with assistance.
4. Search the FAQs at Stata located at <http://www.stata.com/support/faqs/>.
5. Look through the Stata listserv archives at <http://www.stata.com/statalist/archive/>.

Guidelines on Submitting Assignments

Over the course of the semester I will be grading five computer assignments x 20 students, so I can expect around 100 assignments. In order to do the best possible grading job, I need your assignments to be submitted in a very standardized format. For example, when you submit components of the assignment to the course website, name your files using the convention: your last name, a dash, the assignment #, a dash, and the document type. And unless I inform you otherwise, each assignment should consist of the following, and in the order and format listed.

1. Your Stata syntax file used to generate the output. E.g., for the first assignment you would name this *lastname-hw1-syntax.do*, plugging in your last name at the front of the filename.
2. Your Stata output file used to answer the questions. E.g., name this *lastname-hw1-output.txt*.
3. Your answer document one, e.g., named *lastname-hw1-answers.doc*. This is a MS Word document to which you will add brief answers. You can download the blank answer document from Blackboard along with the assignment instructions. If a question in the answer document asks for specific numbers from Stata, copy and paste those numbers directly from your Stata output log file to the answer document. It is VERY important that you keep these numbers in a fixed font and that the lines do not wrap (e.g., Courier, 9 or 10pt font). If your homework is difficult to grade due to formatting problems, I will request that you redo it. In the answer document you'll also describe what you did generally, what data set you used, and how you constructed the measures used in the assignment and dealt with things like missing data.

You should submit all three files to the course Blackboard site by midnight of the due date.

Course Design and Philosophy

SYA 5407 is one of the required methods courses that must be taken and passed by all sociology PhD students before being admitted to candidacy. Why? The most common response to this question is that all sociology PhDs need to have the knowledge/skills necessary to comprehend and evaluate research that is presented at professional sociological meetings (or job talks) and that is published in mainstream social science journals such as *Social Problems*, *Social Forces*, *American Journal of Sociology*, and *American Sociological Review*. RMLDVs occupy a significant and growing portion of the pages that fill these journals. With this central goal in mind, I've designed the course requirements to emphasize understanding of the basic concepts and the ability to evaluate existing work that uses RMLDVs.

There are also secondary goals, and these apply differently across students in the PhD program. Among quantitatively oriented students, it is essential to have an understanding of RMLDVs and be able to use and interpret them appropriately. Further, in this course you will learn many of the tricks that Stata can do to make cleaning and analyzing data and reporting statistical results much easier (this is one reason I am such a big fan of Stata, as you will see). However, it is my strong belief that students with other methodological inclinations will benefit from taking this course as well, above and beyond learning how to read the journals. I think all students can benefit from learning a highly systematic and well-documented approach to social research. "Serious data analysis" (Long & Freese, pp. 28-29) requires that you be able to retrace your steps from a completed research project all the way back to the original data upon which your research is based. These rules for systematic research apply to some extent to all types of social inquiry. Here's how each of the four types of course requirements correspond to my goals for SYA 5407:

Surveys – Surveys are mainly intended for me to get a sense of where students are in terms of their grasp on the current week's materials. E.g., if all students get a particular concept, then I don't need to spend much time on that concept in lecture.

Exemplary Articles – The readings and summaries serve to keep your learning grounded in the reality of how RMLDVs are used and presented in real life. Further, collectively the readings will expose you to some of the more recent quantitative research in esteemed sociology journals.

Quizzes – These are intended to test your on-the-spot ability to know when to apply different types of RMLDVs, know how to interpret the results from them, and know how to evaluate published research that uses RMLDVs. Quizzes provide incentives to keep up with the course readings, and they lead to class discussion of questions that were especially challenging.

Assignments – These are the "meat and potatoes" of SYA 5407 (or the "tempeh and tamari" for you vegetarians). Assignments require that you analyze a unique set of variables that interest you and yield interesting results; that you assess which RMLDV is appropriate for these data if any; that you interpret these results from a given RMLDV; and occasionally that you report the results in a journal format table (a breeze when using Stata).

Final Exam – Though not a favorite form of assessment among most of the students I teach, I've come to firmly believe in the pedagogical value of a well-written, cumulative final exam.

Blackboard Technical Matters

It is your responsibility to confirm that you have successfully submitted an online survey or the three files that make up an assignment. For each of these, after submitting you should look in the grades section of Blackboard to verify your survey or assignment was submitted. It should show an exclamation mark (!) indicating “Needs to be graded” and not the symbol of a lock (🔒) that indicates “In progress” or a dash (-) that indicates Blackboard has no record of you trying to submit anything. If you get the dash symbol, it means you can try again. If you get the lock symbol, you likely will not be able to open the assignment or survey again. In this situation, write Annie or me as soon as possible, then check your email for a reply that reports we’ve cleared the gradebook entry and you can try again. Obviously, you should always keep copies of the files you submit for an assignment. Failing to confirm your submission is not a valid excuse for it being late.

Need Data?

I have many, many suggestions on data sets you might use in your assignments. These include the 1998 Early Childhood Longitudinal Survey (kindergarten cohort), the 2002 Education Longitudinal Study, the 1979 or 1997 National Longitudinal Survey of Youth (two different cohorts), the 2002 National Study of the Changing Workforce, or the 2000 U.S. Census. Other data sets may be appropriate, especially if you are currently analyzing data from them (e.g., Miami Youth Transitions Study or the National Survey of Families and Households). You’ll need a data set that includes categorical/continuous predictors and a variety of outcomes—continuous, categorical, ordinal, and one that’s a relatively rare count (the most difficult one).

Other Information

Honor code: This course will uphold the University Honor Code that is based on the premise that each student makes a commitment to avoid any violation of academic integrity (cheating) and refuses to tolerate violations (report cheating). The FSU Main Campus Student Handbook lists the Academic Honor Policy (http://registrar.fsu.edu/student_handbook/handbook_tal.pdf).

Americans with Disabilities Act: Students with disabilities needing academic accommodation should: (1) register with and provide documentation to the Student Disability Resource Center; and (2) bring a letter to the instructor indicating the need for accommodation and what type. For more information about services for FSU students with disabilities, contact:

Student Disability Resource Center	(850) 644-9506 (voice)
08 Kellum Hall	(850) 644-8504 (TDD)
Florida State University	SDRC@admin.fsu.edu
Tallahassee, FL 32306-4400	

Outline of Course Topics and Due Dates

Required reading: [H] = Hoffman book; [DH] = Dunteman's Sage book

Optional reading: [LF] = Long and Freese's Stata book

Part I. Aug. 27, 29; Sept. 5, 10, 12 [No class on Sept. 3, Labor Day]

Topics: OLS Regression, Probability Distributions, ML Estimation, and Stata

Reading: [H] Chap. 1-2, [DH] Chap. 1-6; optional = [LF] Chap. 1-2

Articles: Call, Sprecher, & Schwartz; Reynolds & Ross; Mirowsky [for fun]

Quiz 1: Sept. 5 or Sept. 10

Assnmt 1: Due Sunday, Sept. 16, by midnight

Part II. Sept. 17, 19, 24, 26; Oct. 1, 3

Topics: Binary Outcomes: Logit & Probit Models, Hyp. Testing, Goodness of Fit

Reading: [H] Chap. 3, [DH] Chap. 7; optional = [LF] Chap. 3-4

Articles: Longmore et al.; Schnittker, Freese, & Powell; Farkas & O'Rand

Quiz 2: Sept. 26 or Oct. 1

Assnmt 2: Due Sunday, Oct. 7, by midnight

Part III. Oct. 8, 10, 15, 17

Topics: Ordinal Outcomes: Ordered Probit and Ordered Logit

Reading: [H] Chap. 4; optional = [LF] Chap. 5

Articles: Kelley-Moore et al.; Keene-Reid & Reynolds; Hill

Quiz 3: Oct. 10 or Oct. 15

Assnmt 3: Due Sunday, Oct. 21, by midnight

Part IV. Oct. 22, 24, 29, 31

Topics: Count Outcomes: Poisson and Negative Binomial Regression

Reading: [H] Chap. 6, [DH] Chap. 8; optional = [LF] Chap. 8

Articles: Rose; Smith & Wiest; Lee et al.

Quiz 4: Oct. 24 or Oct. 29

Assnmt 4: Due Sunday, Nov. 4, by midnight

Part V. Nov. 5, 7, 14, 19, 21 [No class on Nov. 12, Veteran's Day observed]

Topics: Analysis of Durations and Events: Discrete Time & Hazard Models

Reading: [H] Chap. 7, [DH] Chap. 9

Articles: McDonough et al.; Sweeney; DeMaris; Upchurch et al.

Quiz 5: Nov. 14 or Nov. 19

Assnmt 5: Due **Friday**, Nov. 30, by midnight

Part VI. Nov. 26, 28; Dec. 3, 5

Topics: Overview of Other Models: Multinomial Logit, Sample Selection, Etc.

Reading: [H] Chap. 5 + reread [H] Chap. 2, [DH] Chap. 4; optional = [LF] Chap. 6

Articles: Earl et al.; Reynolds; Ross et al.

Final Exam: Due Wednesday, Dec. 12, by midnight

Note: Surveys must be completed by midnight BEFORE each Wednesday (exception for 1st week of class). Due dates for assignments (Sunday by midnight except for Assignment 5) are intended to give you the weekend if need be. I encourage you to finish and submit assignments before the deadline, to avoid delays that are beyond your control and to leave time for questions.

Additional Required Reading

Legend: [OLS] ordinary least squares, [L] logit, [P] probit, [ML] multinomial logit, [OP] ordered probit, [OL] ordered logit, [Po] Poisson, [NBR] negative binomial, [EH] event history or hazard analysis, and [MLM] multilevel modeling

- Call, Vaughn, Susan Sprecher, and Pepper Schwartz. 1995. "The Incidence and Frequency of Marital Sex in a National Sample." *Journal of Marriage and the Family* 57(3):639-52. [OLS]
- DeMaris, Alfred. 2000. "Till Discord Do Us Part: The Role of Physical and Verbal Conflict in Union Disruption." *Journal of Marriage and the Family* 62: 683-692. [EH]
- Earl, Jennifer, Sarah H. Soule, and John D. McCarthy. 2003. "Protest Under Fire? Explaining the Policing of Protest." *American Sociological Review* 68: 581-606. [ML]
- Farkas, Janice L. and Angela O'Rand. 1998. "The Pension Mix for Women in Middle and Late Life: The Changing Employment Relationship." *Social Forces* 76: 1007-1032. [P]
- Hill, Mark E. 2002. "Race of the Interviewer and Perception of Skin Color: Evidence from the Multi-City Study of Urban Inequality." *American Sociological Review* 67: 99-108. [OL]
- Keene-Reid, Jennifer and John Reynolds. 2005. "Gender Differences in the Job Consequences of Work-to-Family Spillover." *Journal of Family Issues* 26(3): 275-299. [OP]
- Kelley-Moore, Jessica A. and John G. Schumacher. 2006. "When Do Older Adults Become 'Disabled'? Social and Health Antecedents of Perceived Disability in a Panel Study of the Oldest Old." *Journal of Health and Social Behavior* 47: 126-141. [OP]
- Lee, Matthew R., William B. Bankston, Timothy C. Hayes, and Shaun A. Thomas. 2007. "Revisiting the Southern Culture of Violence." *The Sociological Quarterly* 48: 253-275. [NBR]
- Longmore, MA, WD Manning, PC Giordano, and JL Rudolph. 2004. "Self-Esteem, Depressive Symptoms, and Adolescents' Sexual Onset." *Social Psychology Quarterly* 67: 279-295. [L]
- Mirowsky, John. 2002. "Parenthood and Health: The Pivotal and Optimal Age at First Birth." *Social Forces* 81: 315-349. [nonlinear regression – just for fun]
- McDonough, Peggy, David R. Williams, James S. House, and Greg J. Duncan. 1999. "Gender and the Socioeconomic Gradient in Mortality." *Journal of Health and Social Behavior* 40: 17-31. [EH]
- Reynolds, Jeremy. 2003. "You Can't Always Get the Hours You Want: Mismatches between Actual and Preferred Work Hours in the U.S." *Social Forces* 81: 1171-1199. [ML]
- Reynolds, John and Catherine Ross. 1998. "Social Stratification and Health: Education's Benefit beyond Economic Status and Social Origins." *Social Problems* 45: 221-247. [OLS]
- Rose, Dina R. 2000. "Social Disorganization and Parochial Control: Religious Institutions and Their Communities." *Sociological Forum* 15: 339-358. [Po]
- Ross, Catherine, John Reynolds, and Karlyn Geis. 2000. "The Contingent Meaning of Neighborhood Stability for Psychological Well-Being." *American Sociological Review* 65: 581-597. [MLM]
- Schnittker, Jason, Jeremy Freese, and Brian Powell. 2003. "Who Are Feminists and What Do They Believe? The Role of Generations." *American Sociological Review* 68: 607-622. [L]
- Smith, Jackie and Dawn Wiest. 2005. "The Uneven Geography of Global Civil Society: National and Global Influences on Transnational Association." *Social Forces* 84: 621-652. [NBR]
- Sweeney, Megan M. 2002. "Two Decades of Family Change: The Shifting Economic Foundations of Marriage." *American Sociological Review* 67: 132-147. [EH]
- Upchurch, Dawn M., Carol S. Aneshensel, Clea A. Sucoff, and Lene Levy-Stroms. 1999. "Neighborhood and Family Contexts of Adolescent Sexual Activity." *Journal of Marriage and the Family* 61: 920-933. [EH]