SOCIOLOGY 6707
INTERMEDIATE DATA ANALYSIS
Winter 2020
Blair Wheaton
Department of Sociology

Time: Tuesday 1-3:30, tutorial Fridays 1-2

Place: Room 240, Dept. of Sociology, 725 Spadina Avenue

TAs: Patricia Louie (patricia.louie@mail.utoronto.ca)

Overview

This course functions as a follow-up to a first graduate level statistics course. The obvious goal is to develop the student's skills as both a producer and a consumer of quantitative findings and data, but the less obvious, but equally important, goal is to develop an understanding of the fit between ideas and models, i.e., how ideas are expressed in models.

The course is intended not just for the specialized student, but for a range of students with different needs, such as:

- increased reading breadth in areas of interest, or in key areas of the discipline or the social sciences in general.
- development of data analytic skills and awareness of available choices in data analysis situations.
- comprehension and application of specific techniques to be used in ongoing and future personal research, dissertation research, and research grants.
- learning a language and a thinking framework that gives access to a wider range of the discipline and the social sciences and thus facilitates the generalization of future audiences in one’s own research.

Themes

- The matching of ideas to their representation in models. A common problem in many areas of research is the lack of fit between the ideas stated in a theory and the way the theory is tested. We emphasize the issue of fit and representation of ideas, as “embodied” in the techniques included in the course.

- Practice in data analysis techniques. The course has been designed and run for a number of years with using exercises on the computer, primarily using SAS but also other programs. These exercises each involve data analysis problems that the student articulates. My role in this part of the course is to be available to students to help with execution of computer problems, and to discuss problems in analyzing and interpreting data. I generally encourage students to help each other with the analytic phases of each project.
Choosing the appropriate technique given the analytical situation and type of data. As the course progresses, an increasingly important issue will be choice of technique, as a function of: type of data, sample size, distributions of variables, nature of underlying concepts (continuous vs. categorical), preferred modes of interpretation, and the nature of the question.

Methods and theory closely linked. Different techniques frame specific forms of theorizing that are not possible outside of those methodological frameworks.

**Topic Details**

**Part 1: Generalizing the Regression Model**

- Interactions and their Interpretation
- Nonlinear Regression (functional forms, splines)
- Logistic Regression (binomial, multinomial, ordinal)
- Regression for Nonnormal Variables (Poisson and Negative Binomial)

We will consider the nature and interpretation of interactions (multiplicative effects of variables), nonlinearity (both functional forms and spline regression), generalizations to categorical outcomes (logistic regression), including both dichotomous and multiple category outcomes, and regression for rare and highly skewed outcomes.

**Part 2: Structural Equation Models**

- The Transition from Equations to Models
- Structural Equation Models: An Introduction
- Structural Equation Models: Testing and Fitting
- Cross-Group Structural Equation Models

We will include a section on structural equation models, including a section on the point of specifying models and process, and a section on the basics of structural equation modeling. We will also consider the testing and fitting of SEM models, and the comparison of models across groups.

**Part 3: Hierarchical and Growth Curve Models.**

- Basic 2-level and 3-level HLM models
- Growth Curve Models
- The Generalized HLM Model (Poisson, Logistic).

These methods address the classic theoretical problem of effects across levels of social reality, specifically targeting the effects of social context on individuals. Social context refers to the effects of any shared context with collective membership. The concept is closely related to idea that layers of social reality can be seen as “nesting units” of increasing size and complexity. Thus, you can study the effects of schools on students, of neighbourhoods on families, of family structure on children, of community on individual opportunities, of social structure on individuals, etc. We will also consider hierarchical models for discrete outcomes. The growth curve model is a direct
extension of the general hierarchical linear model, used to track trajectories of change over lives as a function of time. This technique is especially useful for specifying sources of disparities or inequalities in developmental, social, or other life outcomes over time, with an emphasis on the timing in the life course of the appearance of disparities.

**Part 4: Combined Cross-Section/Time Series Analysis: Fixed and Random Effects Regression for Panel Data**

- Fixed and Random Effects Regression for Panel Data
- Fixed Effects Models for Other Techniques in the Course.

This section considers cross-section / time series models, with an emphasis on fixed-effects and random-effects models and what they do and do not accomplish. Fixed effects models are emphasized, primarily because they claim to take into account a broad class of unmeasured variables left out of the regression which may overlap with the effects of the independent variables in the equation. Fixed effects here stand for stable individual differences in all forms, e.g., biological givens, ascriptive social statuses, and family background. We conclude this section by applying the fixed effects model to techniques discussed earlier in the course, including structural equation models, logistic regression, and Poisson models.

**Part 5: Event History/Survival Models**

- The Discrete-Time Event History Model.
- Grouped Continuous Time Models

The course will consider cases where event history models should be used rather than logistic regression, including the many situations where the timing of an event is as important as its occurrence. Often we study events (marriage, promotion, entry into the labour force, childbearing, etc.) which occur at different times for different people. The event history model takes into account both the occurrence of an event and its timing, while logistic regression can only study the occurrence of the event. We will only consider discrete-time models this year, but these models are very flexible.

**Prerequisites**

The course assumes you know the basics of linear regression, including multiple regression. There will be a voluntary review class for basic regression held in the first week or two. Some of the tutorials on Fridays will provide examples of software used in the course. All of the necessary tools for programming are taught in tutorials, using template programs.

**Required Work**

There is a mid-term test in class, discussed below. Beyond that, you will either do three assignments, coupled with a final exam where you will choose to do one of four questions, or you will do two assignments, coupled with a final exam where you will do three of four questions.

Both the TA and I are available throughout assignment work to answer questions about computer issues and the interpretation of assignment questions. *Both the second and third assignments involve choice of a specific technique from a list of three topics* (structural equation models, hierarchical models, and fixed effects panel regression). *What you choose will determine the timing of your due
The first assignment is standard for everyone with a fixed due date. This assignment considers generalizations of the regression model discussed in the first three weeks.

I expect students to work in groups, formed voluntarily and by mutual agreement among students. This is encouraged for three reasons: 1) to distribute the workload; 2) to encourage collective learning and communication of skills and knowledge among students; and 3) to avoid isolating students with specific computer problems. All grades from these assignments are assigned equally to students within groups. Groups must be from 2 to 3 in size.

There will be a final test on the last day of class. This test is short if you choose to do three assignments, and more involved if you choose to do two assignments. To be specific about how this works, there are three topics after the term test to choose from for assignments:

1. Structural Equation Models
2. Hierarchical Linear Models
3. Fixed Effects Panel Regression

We will discuss event history models as the final topic, but it will not be available for assignments. The final will have four questions to choose from, on these topics:

1. Structural Equation Models
2. Hierarchical Linear Models
3. Fixed Effects Panel Regression
4. Event History Models

Your choices involving assignments and the final test are as follows:

- If you do two assignments, you will write a final answering questions on the other three topics on the final, but not in the area of your assignment #2. As a result, you will not have choice about which questions to answer.
- If you do three assignments, you will be able to write a short final choosing one of the two topics you did not choose for assignments #2 and #3.

The first test is designed to provide a review of notes at a crucial point in the course. I have found in the past that students gain in their ability to understand material in the later phases of the course due to this review of material and consolidation of their knowledge in the middle of the course. All tests are “open-book”, i.e., my notes are allowed.

Weights for Required Work

<table>
<thead>
<tr>
<th>Work</th>
<th>Weight for 3 assignment option</th>
<th>Weights for 2 assignment option</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Assignment #1</td>
<td>20%</td>
<td>25%</td>
</tr>
<tr>
<td>2. Term Test</td>
<td>25%</td>
<td>25%</td>
</tr>
<tr>
<td>3. Assignment #2</td>
<td>20%</td>
<td>25%</td>
</tr>
<tr>
<td>4. Assignment #3</td>
<td>20%</td>
<td>0%</td>
</tr>
<tr>
<td>5. Final Test (in class)</td>
<td>15%</td>
<td>25%</td>
</tr>
</tbody>
</table>
Data Used for Assignments

This course is an overview of a series of techniques. I require that everyone use the same class data for assignments. This requirement is based on my experience with more flexible approaches I used in the past, which led to a number of problems. The most important computer issue in a course such as this is not running a procedure – it is manipulating the data. This is a course in data analysis. As a result, I make your coding part of the issue in grading assignments, because this is where so many of the problems in producing credible findings occur.

I require using class data for these reasons: 1) I need to understand myself the structure of the data, the nature of the sample, and whether variables exist that conform to what you want to do, so that I can give advice during assignments; 2) there are few data sets that can be used for all of the techniques in this course, and we do not want to change data sets across assignments (although it may be necessary for HLM); 3) in general, data has to be at least three-wave panel data, or sufficiently clustered to allow for hierarchical models.

This year we will use the NLSY79, the National Longitudinal Survey of Youth, 1979, and and/or the NLSY79 Children and Young Adults, for assignments, except for HLM. The NLSY79 is one of the most widely used data sets in the world: it is a continuing longitudinal study of youth 14-22 in 1979 in the United States. There were a total of 12,686 individuals in the original sample. A sub-sample of individuals was followed every year until 1994, and then every two years after that, until 2014. The age of the sample at that time varied from 49 to 58 years old. You will be able to set up your own account on the NLSY website and download data directly. The NLSY79 Child/YA data set is a separate study of the children of the NLSY79 cohort. There are 11,521 children that have been followed since birth, with the range in ages covering 0 to 42 years old in 2014. Both data sets can be reviewed at the NLSY web site, including the design, topics, measures, and codebooks. You can download the data directly from this site.

We will also have access to specially extracted and anonymized data from the Toronto Study of Neighbourhoods and Well-Being (2011) for the HLM assignment. If you do this assignment, you must sign a release form promising not to use your results for public presentation in any form or for publication. In fact, you can apply for that approval after the course, if results look promising.

Reading

There is no required reading beyond the set of notes I have developed specifically for this course. The material is intended to be a relatively friendly but rigorous discussion of each technique or type of analysis. The notes are sold on an individual basis, and will be available at Three Cent Copy across the street.

I also hand out a reference list for each topic in the course, printed in a separate document. This list will be uploaded to Quercus, or you can get a copy in class. Under each topic of the reference list, I will include both basic introductions and more complete overviews of each technique, as well as a list of substantive applications of the technique in journals.

Class Schedule

The attached schedule shows the topics covered class-by-class, as well as due dates for all required work.
Web Sites with Basic Mathematical and Statistical Help.

I also strongly encourage use of online sources for learning SAS. The UCLA site for SAS is one of the best and publicly accessible here:
http://www.ats.ucla.edu/stat/sas/

Or on You Tube, an introduction to SAS done by SAS itself:
https://www.youtube.com/watch?v=r1Yy_sYbfy0

Or the introductory online course run by Boston University:
https://support.sas.com/edu/schedules.html?ctry=us&crs=PROG1#s1=1

I will also upload to Blackboard free programming SAS guides you can also download online. These include the “famous” Little SAS Book, a general programming guide:

The Handbook of Statistical Analysis Using SAS:

and the SAS programming skills website at Northwestern:
http://www.kellogg.northwestern.edu/researchcomputing/docs/SAS_Programming_Skills.pdf

PLEASE NOTE: I do require assignments in SAS so that I can give advice and grade them properly.

See below for a more complete list of web sites that provide help for basic math concepts and some of the statistical techniques discussed in this course.

<table>
<thead>
<tr>
<th>Web Sites with Basic Mathematical and Statistical Help.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Algebrahelp.com</td>
</tr>
<tr>
<td><a href="http://www.algebrahelp.com/index.jsp">http://www.algebrahelp.com/index.jsp</a></td>
</tr>
<tr>
<td>Derivatives Defined</td>
</tr>
<tr>
<td><a href="http://web.mit.edu/wwwmath/calculus/ispath/unit02.html">http://web.mit.edu/wwwmath/calculus/ispath/unit02.html</a></td>
</tr>
<tr>
<td>Internet Resources for Math</td>
</tr>
<tr>
<td><a href="http://www.langara.bc.ca/mathstats/resource/onWeb/">http://www.langara.bc.ca/mathstats/resource/onWeb/</a></td>
</tr>
<tr>
<td>Linear Algebra Calculator</td>
</tr>
<tr>
<td><a href="http://www.compute.uwlax.edu/lin_alg/">http://www.compute.uwlax.edu/lin_alg/</a></td>
</tr>
<tr>
<td>Logarithms Definition</td>
</tr>
<tr>
<td><a href="http://www.purplemath.com/modules/logs.htm">http://www.purplemath.com/modules/logs.htm</a></td>
</tr>
<tr>
<td>Logarithms Rules</td>
</tr>
<tr>
<td><a href="http://www.purplemath.com/modules/logrules.htm">http://www.purplemath.com/modules/logrules.htm</a></td>
</tr>
<tr>
<td>Arizona Mathematical Software</td>
</tr>
<tr>
<td><a href="http://math.arizona.edu/~www_main_2002/software/azmath.html">http://math.arizona.edu/~www_main_2002/software/azmath.html</a></td>
</tr>
<tr>
<td>Probability and Statistics</td>
</tr>
<tr>
<td><a href="http://www.ability.org.uk/probstat.html">http://www.ability.org.uk/probstat.html</a></td>
</tr>
<tr>
<td>S.O.S. Math</td>
</tr>
<tr>
<td><a href="http://www.sosmath.com/">http://www.sosmath.com/</a></td>
</tr>
<tr>
<td>Derivatives: Rules and Examples</td>
</tr>
<tr>
<td><a href="http://people.hofstra.edu/Stefan_Waner/RealWorld/tccalcp.html">http://people.hofstra.edu/Stefan_Waner/RealWorld/tccalcp.html</a></td>
</tr>
<tr>
<td>Online Statistical Test</td>
</tr>
<tr>
<td><a href="http://www.stat.ucla.edu/~dinov/courses_students.dir/Applets.dir/Normal_T_Chi2_F_Tables.htm">http://www.stat.ucla.edu/~dinov/courses_students.dir/Applets.dir/Normal_T_Chi2_F_Tables.htm</a></td>
</tr>
</tbody>
</table>
The following parts of the syllabus are a required template, but also express important principles, procedures, and values.

Penalty for Lateness Clause

For both undergraduate and graduate courses, instructors are not obliged to accept late work, except where there are legitimate, documented reasons beyond a student’s control. In such cases, a late penalty is normally not appropriate.

In this course, assignments are only accepted up to two days beyond the due date, and at a 10% discount that applies to the entire group working on the assignment. Late assignments have historically been a rarity in this course for this reason.

Academic Integrity Clause

Copying, plagiarizing, falsifying medical certificates, or other forms of academic misconduct will not be tolerated. Any student caught engaging in such activities will be referred to the Dean’s office for adjudication. Any student abetting or otherwise assisting in such misconduct will also be subject to academic penalties. Students are expected to cite sources in all written work and presentations. See this link for tips for how to use sources well: (http://www.writing.utoronto.ca/advice/using-sources/how-not-to-plagiarize).

According to Section B.I.1.(e) of the Code of Behaviour on Academic Matters it is an offence "to submit, without the knowledge and approval of the instructor to whom it is submitted, any academic work for which credit has previously been obtained or is being sought in another course or program of study in the University or elsewhere."

By enrolling in this course, you agree to abide by the university’s rules regarding academic conduct, as outlined in the Calendar. You are expected to be familiar with the Code of Behaviour on Academic Matters (http://www.artsci.utoronto.ca/osai/The-rules/code/the-code-of-behaviour-on-academic-matters) and Code of Student Conduct (http://www.viceprovoststudents.utoronto.ca/publicationsandpolicies/codeofstudentconduct.htm) which spell out your rights, your duties and provide all the details on grading regulations and academic offences at the University of Toronto.

Accessibility Services

It is the University of Toronto’s goal to create a community that is inclusive of all persons and treats all members of the community in an equitable manner. In creating such a community, the University aims to foster a climate of understanding and mutual respect for the dignity and worth of all persons. Please see the University of Toronto Governing Council “Statement of Commitment Regarding Persons with Disabilities” at http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppno_v012004.pdf.

In working toward this goal, the University will strive to provide support for, and facilitate the accommodation of individuals with disabilities so that all may share the same level of access to opportunities, participate in the full range of activities that the University offers, and achieve their full potential as members of the University community. We take seriously our obligation to make this course as welcoming and accessible as feasible for students with diverse needs. We also understand that disabilities can change over time and will do our best to accommodate you.
Students seeking support must have an intake interview with a disability advisor to discuss their individual needs. In many instances it is easier to arrange certain accommodations with more advance notice, so we strongly encourage you to act as quickly as possible. To schedule a registration appointment with a disability advisor, please visit Accessibility Services at http://www.studentlife.utoronto.ca/as, call at 416-978-8060, or email at: accessibility.services@utoronto.ca. The office is located at 455 Spadina Avenue, 4th Floor, Suite 400.

Additional student resources for distressed or emergency situations can be located at distressedstudent.utoronto.ca; Health & Wellness Centre, 416-978-8030, http://www.studentlife.utoronto.ca/hwc, or Student Crisis Response, 416-946-7111.

**Equity and Diversity**

The University of Toronto is committed to equity and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect. As a course instructor, I will neither condone nor tolerate behaviour that undermines the dignity or self-esteem of any individual in this course and wish to be alerted to any attempt to create an intimidating or hostile environment. It is our collective responsibility to create a space that is inclusive and welcomes discussion. Discrimination, harassment and hate speech will not be tolerated.

Additional information and reports on Equity and Diversity at the University of Toronto is available at http://equity.hrlandequity.utoronto.ca.
<table>
<thead>
<tr>
<th>05</th>
<th>06</th>
<th>07</th>
<th>08</th>
<th>09</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introduction Interactions in Regression</td>
<td>Review: Basic Regression</td>
<td>NO CLASS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear Regression; <em>Downloading NLSY data</em></td>
<td></td>
<td></td>
<td>Intro to SAS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression: Binomial,Multinomial,Continue SAS</td>
<td></td>
<td></td>
<td>NO CLASS</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
</tr>
</thead>
<tbody>
<tr>
<td>Logistic Regression: Ordinal</td>
<td></td>
<td></td>
<td>GLM: Poisson and Negative Binomial Regression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sun</td>
<td>Mon</td>
<td>Tue</td>
<td>Wed</td>
<td>Thu</td>
<td>Fri</td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>01</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>02</strong></td>
<td><strong>03</strong></td>
<td><strong>04</strong></td>
<td><strong>05</strong></td>
<td><strong>06</strong></td>
<td><strong>07</strong></td>
</tr>
</tbody>
</table>
|     |     |     |     |     |     | SEM 1: Equations To Models
| SEM 2: Basic SEM |
| **09** | **10** | **11** | **12** | **13** | **14** | **15** |
| SEM 3: Fitting and Testing Models |     |     |     |     |     | Finish SEM Test Review
| Exercise 1 Due |
| **16** | **17** | **18** | **19** | **20** | **21** | **22** |
| Reading Week | Reading Week | Reading Week | Reading Week | Reading Week |     |     |
| **23** | **24** | **25** | **26** | **27** | **28** | **29** |
|     |     |     |     |     |     | HLM 1: Introduction
<p>| Term Test: 1.5 hours |</p>
<table>
<thead>
<tr>
<th>Date</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>HLM 2:</td>
<td>Examples</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>HLM 3:</td>
<td>Generalized HLM</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>Generalized HLM</td>
<td>Growth Curve Models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>Growth Curve Example</td>
<td>Intro to Panel Models</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td></td>
<td>Panel Regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td>Fixed Effects in SEM</td>
<td>Event History Intro</td>
<td></td>
<td>Exercise 2 Due: HLM</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29</td>
<td></td>
<td>Event History Example</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- SEM Variations Of Fixed Effects Models
- Event History Intro
- Final Review
<table>
<thead>
<tr>
<th>Mon</th>
<th>Tue</th>
<th>Wed</th>
<th>Thu</th>
<th>Fri</th>
<th>Sat</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td><strong>Final Test</strong></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Exercise 2 Due:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panel Regression</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Any Exercise 3 Due:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Any Topic not Ex.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS OF COURSE NOTES

*(RELEVANT PORTIONS IN BLUE)*

**Chapter 1: A Review of correlation and regression**

Overview ................................................................. Error! Bookmark not defined.

1.1 association in A Bivariate Table.................................Error! Bookmark not defined.

1.2 correlation as a measure of association..........................Error! Bookmark not defined.

1.3 Bivariate Regression Theory .....................................Error! Bookmark not defined.

1.4 Partitioning of Variance in Bivariate Regression ..............Error! Bookmark not defined.

1.5 Bivariate regression example .....................................Error! Bookmark not defined.

1.6 assumptions of the regression model............................Error! Bookmark not defined.

1.7 multiple regression...............................................Error! Bookmark not defined.

ConclUding words .....................................................Error! Bookmark not defined.

**Chapter 2. Generalizations of regression 1: testing and interpreting interactions**

Limitations of the Additive Model ..................................... Error! Bookmark not defined.

2.1 Interactions in Multiple Regression.............................Error! Bookmark not defined.

2.2 A Three-Way Interaction between education, race, and gender ......Error! Bookmark not defined.
chapter 3. generalizations of regression 2: nonlinear regression

interactions involving continuous variables

Interpreting an Interaction with Two Continuous Variables

If we work out the net effects of education at two ages—26 and 50—we get the following results, using the substitution of A=26 and then A=50 in the equation above:

Interactions Between Categorical Variables: The N-WAY Analysis of variance

2.4 Interactions Between Categorical Variables: The N-WAY Analysis of variance

2.5 Examples of interactions in published research

2.6 Cautions

The Issue of a “Main Effect” in the Presence of Interactions

Problems with Standardized Solutions

Number of Post-Hoc Tests

2.5 Examples of interactions in published research

2.5.1 The Gender-Specific Effect of Marriage

2.5.2 Two Distinct Issues in an Interaction: Race, Gender, and Chains of Disadvantage

2.5.3 A Three-Way Interaction

Concluding words

chapter 3. generalizations of regression 2: nonlinear regression

3.1 A simple example of a quadratic relationship

3.2 estimating higher order relationships

3.3 A Digression: Basic Math for nonlinear models

3.4 interpretation of nonlinear functions

3.4.1 Interpreting Powers of X

Quadratic Curve

Cubic Curve

3.4.2 Log Transformations

5.1 The POISSON REGRESSION model .................................................. Error! Bookmark not defined.
   Negative Binomial Model ................................................................. Error! Bookmark not defined.
   5.1.1 The Number of Physical Arguments per Year in Married Couples........ Error! Bookmark not defined.
   An Example of Poisson Regression ................................................. Error! Bookmark not defined.
   Differences in Results across Models ............................................ Error! Bookmark not defined.
   5.1.2 Interpreting the Results of the Poisson / Negative Binomial Model—Three Ways:.... Error! Bookmark not defined.

5.2 The Complementary log-log model .................................................. Error! Bookmark not defined.
   5.2.1 An Example of the Complementary Log-log Model: Being a Victim of Domestic Violence .......... Error! Bookmark not defined.
   5.2.2 Interpreting the Results of the Complementary Log-log Model:...... Error! Bookmark not defined.
   The Relative Risk vs. the Odds Ratio ............................................... Error! Bookmark not defined.

5.3 examples in published research ..................................................... Error! Bookmark not defined.
   5.3.1 Protest Behavior .................................................................. Error! Bookmark not defined.
   5.3.2 School Violence ................................................................. Error! Bookmark not defined.

concluding words.............................................................................. Error! Bookmark not defined.

6.1 What is wrong with equations?......................................................... Error! Bookmark not defined.
   6.1.1 Regression as a Competition among Predictors .............................. Error! Bookmark not defined.

6.2 Equations vs. Models: Some Examples ............................................ Error! Bookmark not defined.
   6.2.1 The Difference Between Direct Effects and Total Effects ............. Error! Bookmark not defined.
   6.2.2 When A “Control” Variable is not a Control Variable .................... Error! Bookmark not defined.

6.3 Why Causality?.............................................................................. Error! Bookmark not defined.
   6.3.1 Causal Claims and Causal Evidence ........................................ Error! Bookmark not defined.
   6.3.2 Causality vs. Association ...................................................... Error! Bookmark not defined.
Chapter 7. An Introduction to Structural Equation Models

6.4 criteria for causality
   6.4.1 Criteria for the Social Sciences
   Excluded Criteria

6.5 The analytical roles of Variables in causal models
   6.5.1 Focal Association
   6.5.2 Roles of Other Variables
   6.5.3 Assigning Roles to Variables in an Analysis
   Distinguishing Controls from Mediators

6.6 Interpreting an association using controls and mediators
   6.6.1 The Structural Equations in Standardized Form
   6.6.2 A Scheme for Decomposing Correlations into Parts
   Types of Direct and Indirect Effects
   6.6.3 Estimating the Model
   6.6.4 An Expanded Model

6.7 Special Cases
   6.7.1 Suppressor Variables
   6.7.2 Two Examples
   Time Series of Suicide Rates
   The Black-White Paradox in Mental Disorder
   6.7.3 Mediating / “Moderating”

6.8 From Recursive to Non-Recursive Models: What to do about reciprocal Causation
   Terminology for Structural Equation Models
   Notation for Structural Equation Models
   The Consequences of Reciprocal Causation
   6.8.1 The Identification Problem
   T-Rule for a Model: (a necessary condition)
   A Rule of Thumb for Identification
   6.8.2 Identification of Recursive Models

6.9 further examples
   6.9.1 Using a Model to Decide Theory: The Miller-Stokes Model of Democratic Representation
   Impact of Spousal Work-Family Conflict

concluding words

Chapter 7. An Introduction to Structural Equation Models
Chapter 9. Variations and Extensions of SEM

7.2.1 The Exploratory Factor Analysis Model
Generalized T-Rule for Identification
7.2.2 The Confirmatory Factor Model
7.2.4 Back to the Causal Model for Observed Variables
7.3 The FULL SEM model
7.3.1 A Brief Note on Estimation
Maximum Likelihood (ML) Fitting Function
Goodness of Fit
7.3.2 Example 1: A Confirmatory Factor Model
7.3.3 Example 2: Social Class, Mastery, and Depression
Estimating the Model in SAS
7.3.4 Selected Output
Modeling Information
Descriptive Statistics
Covariances
Fit Statistics
Unstandardized Output
Standardized Output
7.3.4 Example 3: Income and Depression
7.3.5 Example 4: The Reciprocal effects of Public and Private Life Satisfaction

7.4 examples in published research
7.4.1 Gender Role Attitudes and Marital Quality
7.4.2 Community Disorder, Victimization, and Mental Health among Youth

concluding words

Chapter 8. IDENTIFICATION AND TESTING OF MODELS
8.1 identification
8.1.1 A Three -Stage Strategy
8.1.2 The General T-Rule
8.1.3 Identification of Measurement Models
Applying the FC1 Rule
8.1.4 Identification of Structural Model
The Distinction between Recursive and Nonrecursive Models
Recursive Models
Nonrecursive Models
Nonrecursive Model with Reciprocal Effects and/or Correlated Error
Order Condition
The Rank Condition
Distinguishing Between Just-identified and Over-identified Equations

8.2 testing and fitting models
2.1 Four Measures of Overall Fit Status
8.2.2 Measuring the Incremental Change in Fit
The “One Step Beyond” Rule

8.3 examples of fitting models
8.3.1 The Mastery Model
8.3.2: The Relationship between Public and Private Life Satisfaction

concluding words

Chapter 9. Variations and Extensions of SEM
9.1 the comparative SEM framework
  9.1.1 Implementing Multiple-Group Models
  9.1.2 Partitioning of Variance and the Role of Contextual Differences
  9.1.3 Developing a Model at Each Level
  9.1.4.1 The Structure of the Model
  9.1.4.2 The Level
  9.1.4.3 One way ANOVA with Random Effects
  9.1.4.4 Fitting the Model First
  9.1.4.5 Rules for Selecting Instrumental Variables
  9.1.4.6 Using Marginal Maximum Likelihood Estimation in a NonLinear Mixed Model

9.2 A Multiple Group Example
  9.2.1 The SES-Depression Model among Black and White Americans
  9.2.2 Estimating the Model
  9.2.3 The Full “Mixed” Model
  9.2.4 T

concluding words
10.6 sample size considerations

10.7 estimating multilevel models IN SAS

10.7.1 Neighborhood Effects On Immigrant Women's Alcohol Consumption

10.7.2 General Background to Preparation of Data

10.9 further Examples from the literature

10.9.1 Understanding Spatial Variations in Tolerance

10.9.2 Contextual Effects in the Darfur Genocide

10.9.3 The Effects of Classroom Context on First Grade Children

concluding words

Chapter 11. the generalized hierarchical Linear Model

11.1 multilevel logistic regression

11.1.1 Estimation

11.1.2 Generalization to Other Logistic Models

11.1.3 Why Use a Hierarchical Approach?...an Example

11.2 Running the generalized HLM in SAS

11.3 multilevel poisson regression

11.4 a published 2-level logistic example

concluding words

Chapter 12. GROWTH CURVE MODELS

12.1 deriving the structure of growth models

12.1.1 Individual Trajectories

12.1.2 The Effect of Early Instruction on Cognitive Growth

12.1.3 A Quadratic Growth Model

12.1.4 Piecewise Linear Growth Models

12.2 running growth models in SAS

12.2.1 Data Structures for Growth Models

12.3 modeling the trajectory of net worth From early to mid adulthood

12.3.1 Setting Up the Data

12.3.2 The Unconditional Growth Model

12.3.3 Adding the Effect of Marital Status and Marital Transitions

12.3.4 The Effects of Childhood Poverty on Initial Status

12.3.5 The Impact of Poverty on Net Worth Over Time

12.4 Modeling the Trajectory of Internalizing Problems over Adolescence

12.5 published examples

12.5.1 Gender and Race Differences in Occupational Stratification over the Life Course

12.5.2 Social Dynamics of Health Inequalities
Chapter 13. introduction to regression for panel data

13.1 The generalized panel regression model
13.1.1 Panel Data Sets
13.1.2 The General Model
13.1.3 The “Unrestricted” Model: The Variable Intercepts/Variable Slopes Model
13.1.4 The Variable Intercepts/Constant Slopes Model
13.1.4.1 Fixed Effects Model
13.1.4.2 Random Effects Model
13.1.5 The Constant Intercept/Constant Slopes Model
13.2 examples of panel regression
13.2.1 The Effects of Marital Status on Depression: 1988-2002
13.2.1.1 Creating Person-Time Data
13.2.1.2 Running PROC PANEL in SAS
13.2.1.3 Models to Test the Stability of Effects over Time
13.2.1.4 Interpreting Effects in a Panel Regression
13.2.2 The Effects of SES (Education and Income) on the Body Mass Index
13.3 Applying Fixed Effects in Published research
13.3.1 The Effects of State Beer Taxes on Automobile Fatality Rates
13.3.2 Explaining Occupational Sex Segregation and Wages
13.3.3 The Effects of Class Size on Public University Student Grades

CONCLUDING words

Chapter 14 Variations and extensions of panel regression

14.1 Models for the Effects of Changes between Waves
14.1.1 Modeling Change Events in Multi-Wave Data
14.1.1.1 A Caveat: Two Wave Data
14.1.1.2 Multi-Wave Data
14.1.2 Evaluating the Time-Dependent Effects of Change
14.1.2.1 The Effects of Time on the Consequences of Divorce and Widowhood
14.1.2.2 The Effect of Divorcing at:
14.1.2.3 The Effect of Counteracting Events
14.1.2.4 The Effects of Marriage on BMI among Women
14.2 Dynamic Panel Models
14.3 fixed effect methods for logistic regression
14.4 Fixed-Effects methods for structural equation models
14.4.1 A Basic Model
14.4.2 A General Model for Causal Inference in Panel Data
14.4.2.1 Fixed Effects in Estimating Effects in Women’s Work Careers
14.4.2.2 LINEQS Statement
14.4.2.3 STD and COV Statements

A Published Example: The Effects of Parental Loss Using fixed effects
Amato and Anthony (2014): The Effects of Parental Divorce and Death

CONCLUDING words
Chapter 15. Event History Analysis in discrete time

15.1 Overview of concepts and models
15.1.1 Problems with Traditional Approaches
15.1.2 What is an Event History and What Does it Do?
15.1.3 The Risk Period
15.1.4 Two Models: Discrete vs. Continuous Time

15.2 the Discrete-time event history model
15.2.1 Introduction

15.3 Basic Concepts
15.3.1 The Survival Function
15.3.2 The Hazard Function
15.3.3 Differences in Hazard Functions

15.4 Creating and analyzing A Person-Period Data set
15.4.1 Form Of The Data
15.4.2 Results for a Model with Four Main Effects
15.4.3 Other Issues
15.4.3.1 Interactions with Time and Nonproportional Hazards
15.4.3.2 Modeling Time
15.4.3.3 Competing Risks
15.4.3.4 Repeated events

15.5 studying Women’s entry into the Work role after Having a First Child

15.5.1 Defining the At-Risk Population
15.5.2 Preliminary Variables Needed To Define Onset of Risk
15.5.3 Focal Variable in Analysis
15.5.4 Developing Measures of Work History to Detect the Timing of the First Job after the Birth of the First Child
15.5.5 Information Used to Define Censoring
15.5.6 Developing Indicators of Censoring and Duration of Risk
15.5.7 Fixed Control Variables
15.5.8 Time-Varying Variables
15.5.9 Checking Coding
15.5.10: Creating the Person-Period Data Set and Collapsing Time Periods
15.5.11: Checking the Person-Period Data
15.5.12: Analyzing the Person-Period Data
15.5.12.1 The Time-Only Model
15.5.12.2: The Effect of Decade of Birth
15.5.12.3: Adding Control Variables
15.5.12.4: Testing the Interaction between Time Period and Decade

A PUBLished Example: Combining event history with HLM Models

15.6 CONCLUDING words

Chapter 16: The continuous time event history model

16.1 Hazard and Survival Concepts in Continuous Time
16.1.1 Hazard and Probabilities in Continuous Time
16.1.2 The Risk Set
16.1.3 The Risk Population

16.2 The Proportional Hazards Model