MULTIVARIATE ANALYSIS
SYA 5406 Section 1
Spring 2015
Tues/Thurs 12:30PM-1:45PM
RBB323

Instructor  Koji Ueno  Teaching Assistant  Kristen Erichsen
Email  kueno@fsu.edu  Email  ke15@my.fsu.edu
Office  Bellamy 512  Office  Bellamy 506
Phone  850-645-2438  Office Hours  By appointment

COURSE DESCRIPTION
This course covers basics of multivariate analysis. We will first review bivariate statistics and the logic of statistical inference and move on to multivariate statistics focusing on ordinary least square (OLS) models. The course is designed to increase understanding of the assumptions of OLS models and problems that result from violations of these assumptions (e.g., multicollinearity, autocorrelation, heteroscedasticity) as well as techniques to reduce these problems (e.g., variable transformation, alternative models). A segment of the course is devoted to techniques of testing different causal processes such as moderation and mediation. In addition to these conceptual and mathematical issues, the course is designed to help students develop practical skills for managing and analyzing secondary data using a software package called Stata.

PREREQUISITE
Successful completion of SYA5305 (Introduction to Research Methods) is required. It is important that you have mastered basic concepts in statistics (sampling distribution, hypothesis testing, estimation/confidence interval, bivariate analysis such as simple regression). Prior experience with Stata is not required, but you should have some previous exposure to programming in statistical software (as opposed to clicking options). If you would like to use other graduate statistics courses to waive the prerequisite, please come talk to me.

COURSE OBJECTIVES
Upon successful completion of this course, you will master the following skills:

1. Explain the mathematical background of OLS and related models (e.g., how coefficient estimates are calculated).
2. Identify an appropriate model for a given research question and a data set.
3. Run multivariate models using Stata syntax.
4. Interpret Stata output.
5. Evaluate whether a given data set meets the OLS assumptions.
6. Determine what procedures need to be taken for a given type of OLS assumption violation.

In addition to these objectives specific to the course content, you are expected to develop and improve the following general skills in statistics:

1. Learn new statistical techniques. Although the basic concepts of multivariate analysis will probably remain important, some statistical techniques will become obsolete, and new techniques will be available over time. Your future research projects may require you to learn specialized statistical techniques. To prepare for these possibilities, you are expected to develop general skills to update and expand your statistical knowledge, as you learn the course materials.
2. Learn new or different statistical software packages. As with conceptual skills, your computing skills will require updates and improvements. As you learn Stata, you will develop general skills to read manuals, use built-in “help” functions, locate appropriate texts, contact software companies, gather information from the internet, etc.
3. Identify statistical techniques most relevant to your research areas and start developing expertise in them. Depending on your research interests, some multivariate models are more important than others. Although our main goal in this course is to develop a broad understanding of multivariate models, you should begin to think about what statistical techniques you want to study in more depth in the future.
4. Learn how to obtain existing data and prepare them for analysis. There are many secondary data sets publicly available and others for which you can obtain a contract. Through your research project (see
below), you will learn how you can put together secondary data, convert them in an appropriate format, construct variables, and so on.

**BASIC FLOW OF THE COURSE**
For each chapter/topic:

- **Home**: Read
- **Class**: - Concepts & Math
- **Review**: - Chapter
- **Next Topic**: Exercise Feedback

Simultaneously, you will work on your research project.

**COURSE MATERIALS**

**Required Text:**

*This textbook covers most of the important topics without presenting too much information. Please consider this book as a starting point and use other texts as supplements (see the “recommended texts” section below).*

**Other Required Readings:**
Will be posted on Blackboard.

**Optional Text:**

*This will be our reference book for Stata. I strongly encourage you to purchase a copy unless you are already a Stata expert. Some of you may have Alan Acock’s A Gentle Introduction to Stata. There is substantial overlap between these two texts, so if you have Acock’s book, use it as your Stata reference, instead of Hamilton’s book.*

**Stata:**
Our sociology graduate lab has six computers with Stata SE (Special Edition) on them. If you would like to purchase a copy for your laptop/home computer, you have the following choices: perpetual license of Special Edition ($395) or Intercooled Edition ($235) or one-year license of Special Edition ($198) or Intercooled Edition ($125). Go to [http://www.stata.com/order/schoollist.html](http://www.stata.com/order/schoollist.html) to learn more and order your copy. After you place an order, you can pick up your installation CD and manual from the Stata contact person at FSU.

*The difference between SE and Intercooled:* SE allows datasets with up to 32,767 variables and can fit models with more independent variables (up to 10,998). Intercooled version allows for datasets with up to 2,047 variables and models with 798 independent variable variables at most. For both versions, the number of observations is limited by the amount of RAM in your computer.

*Non-sociology students:* Use of Sociology Lab is limited to sociology students. Please contact your department to see whether you have access to Stata in your department. If not, I recommend you buy your own copy.
COURSE REQUIREMENTS

Attendance:
Class attendance is required. Attendance is an essential element of professionalism, which you are expected to develop as a graduate student. Attendance will not directly influence your final grade, but I keep attendance records.

Policy for cell phones and handheld devices: I ask you to keep your cell phones and handheld devices turned off during class sessions. If you need to receive emergency calls, please use the vibrate setting.

Policy for laptop use in class: You are allowed to take notes using a laptop computer. Students who use their computers in any other ways may be asked to leave the classroom.

Reading Assignments:
Please read the assigned chapters before you come to each session. Keep in mind that you are learning to become an independent researcher, so read the chapters as if you are trying to learn the materials on your own. You may find some chapters difficult to follow, but don’t worry. Just try to understand the big picture and ask any questions in class.

Chapter Exercises:
I will select a few problems from each text chapter. An assignment sheet will be posted on Blackboard after we go over each chapter in class. I ask you to submit the following two items for each assignment on Blackboard:

1. A main document that describes your answers. Explain in words most important information from your output. Just saying “see output” is not sufficient in most cases. Please type your document in a word processing program such as Microsoft Word. Feel free to paste important parts of the output including graphs.
2. A Stata log file. You do not have to submit graph files (e.g., histograms, scatter plots) unless the assignment sheet asks you to do so.

Collaboration policy:
I encourage you to help each other by exchanging notes, going over lecture notes, discussing each other’s research project, and so on. To ensure that each of you will develop competency as an independent researcher, however, I ask you to first work on assignments individually and collaborate later. Here are some examples of questions you can ask to each other:

- “My answer to Question 2 is 3.5. How about yours?”
- “I keep getting error messages in Stata. Can you look at my syntax and output?”
- “I am not sure whether I am on the right track or not. Let me explain how I am trying to interpret the output, and can you tell me what you think?”
- “I don’t know how I can get started on Question 3. Can you tell me which part of the lecture notes and textbook I should look at?”

The following types of questions are not acceptable:

- “Can you show me the Stata command you used?”
- “I don’t have time to answer your question, can I just show you the Stata command I used?”
- “Let’s work on this problem together.”
- “I forgot to save my Stata output. Can you send me yours?”

Time management suggestion:
Statistical analysis can be stressful because one simple error in syntax could prevent you from producing correct output. Now imagine you are trying to complete your assignment in the last minute! You have a week to complete each assignment. I encourage you to produce tentative Stata output within a couple of days each assignment is posted. That way, you can ask any questions in class and still have a few more days to correct any errors and write up your report.
Getting help:
If you have questions or need help completing chapter exercises, please take the following steps:
1. Ask questions in the next class session.
2. Send your question to the class listserv (mvsoc@fsu.edu). Instead of sending me individual emails about chapter exercises, please use the listserv. I also encourage you to answer each other’s questions. You can also use the listserv to ask questions regarding other issues (e.g., general questions about research projects, Stata, class schedule). You will be added to the listserv in the first week of semester.
3. Come to my office hours or make appointments with the teaching assistant. If my office hours do not work for you, just send me email to set up a time.
4. If necessary, we will arrange weekly lab sessions.

When getting help, please be prepared to present the following things:
- When asking about error messages in Stata, tell us what strategies you have tried to solve the problem.
- When asking about output interpretations, tell us which part of textbook and lecture notes you consulted and why they did not help your interpretations.

If you find yourself needing help in every step, you probably don’t have a good understanding of the chapter. Please review the textbook and lecture notes carefully before starting to work on the assignment again.

Research Project:
By completing a series of assignments, you will engage in a small research project. The purpose of this assignment is to experience the whole process of secondary data analysis—getting data, reading codebooks, constructing variables, running models, interpreting output, revising models, and writing up the results. Assignment sheets will be posted on Blackboard to provide instructions for each step, and you will submit your work on Blackboard.

Exams:
You will take two in-class, closed-book exams (midterm and final). Each test consists of short-answer questions. The final exam is not cumulative. Details will be announced later in the course.

GRADING
Chapter exercises (7) and Individual Project Tasks (5):
For each assignment, you will receive a √ (“check”), √+, or √-. To compute your final assignment grade, each assignment grade will be converted into a 10-point scale as follows: check plus (10 points), check (9.5 points), or check minus (8.5 points). A “check” indicates you demonstrated basic understanding of the materials, and a “check minus” is given when the assignment contains incorrect interpretations, major programming errors in Stata, or general sloppiness. “Check pluses” are reserved for exceptional work. If you receive an assignment back without a grade, you will need to redo it. If the resubmitted assignment is satisfactory, you will receive a check minus.

Late Assignment Policy:
Chapter exercises and project assignments are due on Blackboard when the class session begins at 12:30PM. Assignments are designed to help you master each topic before moving to the next one, so late submission will miss the purpose of these assignments. You have at least a week to complete the assignment, so technical problems in the last minute (e.g., Blackboard server dysfunction) are not an excuse for late submission. To encourage timely submission, one point will be deducted from assignments turned in late on the due day or on the next day. If you were unable to submit assignments after the next day, please contact me right away.

Exams (2):
100 points each.

Computing the Final Grade:
Your final grade will be computed as follows:
- Chapter exercises: 20%
- Individual Project: 15%
- Midterm Exam: 30%
- Final Exam: 35%
The final score will be converted to a letter grade as follows:

<p>| | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>A+</td>
<td>A</td>
<td>A-</td>
<td>B+</td>
<td>B</td>
<td>B-</td>
</tr>
<tr>
<td>93-100</td>
<td></td>
<td>89-92</td>
<td>86-88</td>
<td>83-87</td>
<td>80-82</td>
<td>77-79</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>C+</td>
<td>C</td>
<td>C-</td>
<td>D+</td>
<td>D</td>
<td>D-</td>
</tr>
<tr>
<td>79-82</td>
<td>76-78</td>
<td>73-75</td>
<td>70-72</td>
<td>67-69</td>
<td>64-66</td>
<td>60-62</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>50-59</td>
<td>46-49</td>
<td>42-45</td>
<td>39-41</td>
<td>35-37</td>
<td>30-32</td>
</tr>
</tbody>
</table>

### COURSE WEB SITE

When you register for this course, you automatically gain access to the course web site (Blackboard). In this web site, you can download assignment sheets, submit your work, view your assignment/test scores, etc. To visit the site, open your web browser and go to [https://campus.fsu.edu](https://campus.fsu.edu). After you log on, click the course number. If you are not familiar with Blackboard, click on “Blackboard Tools” in the first logon page to learn more.

### TEXTBOOK WEB SITE

Our textbook (CH) has a web site: [http://www1.aucegypt.edu/faculty/hadi/RABE5/index.html](http://www1.aucegypt.edu/faculty/hadi/RABE5/index.html). From the web page, you can download data sets used in the textbook. The site also includes errata for the textbook. You may want to make corrections in your copy using the information.

### OTHER RESOURCES

#### Recommended Texts:

**Monographs from Sage “Quantitative Applications in the Social Sciences”:**

*Also known as “green books.” Each volume covers a specific topic. Among many volumes available, the following ones are most relevant to our course:*

- Hardy, Melissa. *Regression with Dummy Variables*. No. 93
- Pampel, Fred C. *Logistic Regression*. No. 132.
- Dunteman, George H. *Principal Components Analysis*. No. 69.
- Spector, Paul E. *Summed Rating Scale Construction*. No. 82.

#### General Texts:

*You may find the following books helpful, if CH does not fit your style or background knowledge, or if you would like to have multiple sources of information.*


#### Stata Books and Resources (in the order I recommend):

Built-in “Help” function in Stata.

*From the top menu, click “Help” ➔ “Search”.*

Stata Base Reference Manuals.

*From the top menu, click “Help” → “PDF Documentation”. The manuals provide mathematical background as well as information about how to specify subcommands and options for each procedure. If you do not get enough information from the built-in help, these manuals may be useful.*


Email to Stata. Read instructions at [http://www.stata.com/support/tech-support/](http://www.stata.com/support/tech-support/).


**ACADEMIC HONOR CODE**

You are expected to uphold the Academic Honor Code published in The Florida State University Bulletin and the Student Handbook. The Academic Honor System of The Florida State University is based on the premise that each student has the responsibility (1) to uphold the highest standards of academic integrity in the student's own work, (2) to refuse to tolerate violations of academic integrity in the university community, and (3) to foster a high sense of integrity and social responsibility on the part of the university community.

Please see the following web site for a complete explanation of the Academic Honor Code.

[http://fda.fsu.edu/Academics/Academic-Honor-Policy](http://fda.fsu.edu/Academics/Academic-Honor-Policy)

**DISABILITY ACCOMMODATIONS**

Students with disabilities needing academic accommodation should: (1) register with and provide documentation to the Student Disability Resource Center; (2) bring a letter to the instructor indicating the need for accommodation and what type. This should be done during the first week of semester.

For more information about services available to FSU students with disabilities, contact:

- Student Disability Resource Center
- 97 Woodward Avenue South, Florida State University, Tallahassee, FL 32306-4167
- Voice: (850) 644-9566; TDD: (850) 644-8504; Email: [sdrc@admin.fsu.edu](mailto:sdrc@admin.fsu.edu);
- Web page: [http://www.fsu.edu/~staffair/dean/StudentDisability/](http://www.fsu.edu/~staffair/dean/StudentDisability/)

**SYLLABUS CHANGE POLICY**

This syllabus is a guide for the course and is subject to change with advanced notice.
# COURSE CONTENT AND SCHEDULE

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Topic</th>
<th>Assignment Due</th>
</tr>
</thead>
<tbody>
<tr>
<td>January 10</td>
<td>Welcome; Getting Data</td>
<td></td>
</tr>
<tr>
<td>January 12</td>
<td>(Review HM2-5 or AC2-9) Getting Started with Stata; Univariate and Bivariate Analysis</td>
<td></td>
</tr>
<tr>
<td>January 17</td>
<td>CH1&amp;2 Simple Linear Regression; Putting Data Sets Together</td>
<td>P1</td>
</tr>
<tr>
<td>January 19</td>
<td></td>
<td></td>
</tr>
<tr>
<td>January 24</td>
<td>CH3; (HM6 or AC10) Multiple Linear Regression</td>
<td>Uni/bivariate</td>
</tr>
<tr>
<td>January 26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>January 31</td>
<td>CH4; (HM7,9) Regression Diagnostics</td>
<td></td>
</tr>
<tr>
<td>February 02</td>
<td></td>
<td>CH2,3</td>
</tr>
<tr>
<td>February 07</td>
<td>CH5 Dummy Variables and Interactions</td>
<td></td>
</tr>
<tr>
<td>February 09</td>
<td></td>
<td>CH4</td>
</tr>
<tr>
<td>February 14</td>
<td>(CH11); Aneshensel 4&amp;7 More on Interactions (in-class exercise)</td>
<td></td>
</tr>
<tr>
<td>February 16</td>
<td>CH2,3 Variable Selection Procedures; Theory Based Model Building</td>
<td>CH5</td>
</tr>
<tr>
<td>February 21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>February 23</td>
<td></td>
<td>Catch-up/Q&amp;A</td>
</tr>
<tr>
<td>February 28</td>
<td></td>
<td>Mid Term Exam</td>
</tr>
<tr>
<td>March 02</td>
<td>CH8; (HM13) Autocorrelation; Time Series</td>
<td></td>
</tr>
<tr>
<td>March 07</td>
<td></td>
<td>P3-descriptives</td>
</tr>
<tr>
<td>March 09</td>
<td>CH6; (HM8) Transformation of Variables</td>
<td></td>
</tr>
<tr>
<td>March 14</td>
<td></td>
<td>No Class (Spring Break)</td>
</tr>
<tr>
<td>March 16</td>
<td></td>
<td>No Class (Spring Break)</td>
</tr>
<tr>
<td>March 21</td>
<td></td>
<td>More on Transformation of Variables</td>
</tr>
<tr>
<td>March 23</td>
<td>CH9; (HM12) Multicollinearity, Principal Components</td>
<td></td>
</tr>
<tr>
<td>March 28</td>
<td>CH10; Online</td>
<td>CH6</td>
</tr>
<tr>
<td>March 30</td>
<td></td>
<td>No Class (Southern Sociological Society meeting)</td>
</tr>
<tr>
<td>April 04</td>
<td></td>
<td>More on Multicollinearity, Principal Components</td>
</tr>
<tr>
<td>April 06</td>
<td>CH12; (HM10 or AC11) Logistic Regression</td>
<td></td>
</tr>
<tr>
<td>April 11</td>
<td></td>
<td>Principal Comp.</td>
</tr>
<tr>
<td>April 13</td>
<td>Online Missing Data</td>
<td></td>
</tr>
<tr>
<td>April 18</td>
<td></td>
<td>Student Presentations</td>
</tr>
<tr>
<td>April 20</td>
<td></td>
<td>CH12(optional)</td>
</tr>
<tr>
<td>April 25</td>
<td></td>
<td>Catch-up/Q&amp;A</td>
</tr>
<tr>
<td>April 27</td>
<td></td>
<td>Final Exam</td>
</tr>
</tbody>
</table>

**CH -- Chatterjee & Hadi**  **P1-P5 -- Individual Project Assignments**  
**HM -- Hamilton**  **Chapters in parentheses -- Skim the content**  
**AC -- Acock**