

SOCI 461: Quantitative Data Analysis

Winter 2014
Monday and Wednesday
10:05 p.m. – 11:25 p.m.
Leacock 212

Prof. Aniruddha (Bobby) Das
E-mail: Please see communication policy below
Office Hours: Mon/Wed 4-5 pm.
Office Address: Room 730 Leacock Building

Communication policy

Please use MyCourses for all e-mail communications, keeping the original subject line intact [[Winter 2013 - SOCI-461-001 - Quantitative Data Analysis](#)]: Append your topic after the colon. E-mails sent directly to the McGill general e-mail address will not be answered. I will make every attempt to answer e-mail in a timely fashion within 36 hours of receipt. Please see me during office hours for urgent issues.

Overview

This is the second in a two-course sequence in social statistics. The first (SOCI 350 for Sociology majors) is a prerequisite for this course. SOCI 350 deals with the fundamentals of statistics—including descriptive statistics, hypothesis testing, and the analysis of data involving a few variables (usually one dependent variable and one independent variable). SOCI 461 is a writing intensive, computer-based course dealing with multivariate quantitative data. Because in real-life applications the dependent variable is usually limited or categorical, this course teaches students how to analyze and interpret results from logit, probit, ordered logit, and multinomial logit regression models. The principal objective is to have students participate actively in sociological inquiry and to learn about data analyses by conducting primary research of their own. At the end of the course, students will be able to properly apply different types of regression models and interpret results correctly. These goals will be accomplished through **3 data analyses assignments**, a **team research paper** that is due the last day of class, and in-class **team presentations** of this research project. The dataset to be used is the Chinese Health and Family Life Survey (<http://popcenter.uchicago.edu/data/chfls.shtml>).

Given the emphasis on active learning, much of the class time will be devoted to hands-on data analysis. Students are advised to work with the variables they plan to use for their final paper, during lab sessions and in each of their take-home assignments.

Readings

There are no required readings for this course. All relevant material will be covered in class.

Suggested readings:

Beck-Lewis, Michael S. 1980. *Applied Regression: An Introduction*, vol. 22, Edited by J. L. Sullivan. Newbury Park, California: Sage. (Beck in course outline below)

Baglione, Lisa A. 2007. *Writing a research paper in political science : a practical guide to inquiry, structure, and methods*. Belmont, CA: Thomson Higher Education.

Long, Scott J. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Vol. 7. Edited by Richard Berk. Thousand Oaks, California: Sage. (RMCLDV in outline below)

Long, Scott J. and Jeremy Freese. 2003. *Regression Models for Categorical Dependent Variables Using Stata*, Revised Edition. College Station, Texas: Stata Press. (L & F in outline below)

Evaluation

Three take-home assignments	20*3 = 60%
Team research paper	30%
Paper presentation	10%

Course requirements

Apart from classroom discussions, the course page on MyCourses also has a Discussion Board. You are strongly encouraged to post questions and comments on it, and to respond to existing ones.

There will be approximately one **assignment** every month. The point of the assignments is to help you master the tools for the research paper. To do well in the assignments, keep the following in mind:

- For assignments that ask you to analyze real data, find a topic that shows something interesting. If you pick variables that are not significant and/or have small effects, your grade will be lower.
- All Tables, listings of computer commands, and listings of output must be printed with a fixed font. You must make the font small enough so that a line of output does not wrap from one line to the next. (Set page margins to .5 inches all around, and use Courier New 11 pt font).

You are encouraged to discuss your work with other students. You must complete assignments on your own. On computer assignments, everyone must use their own variables for their analysis. All hands-on work will use Stata.

In addition to the 3 assignments, each student will be assigned to a group for the **team research paper**. (Volunteering for a group is acceptable and encouraged.) Final papers should be about 20 pages (plus references, tables, and figures). Specifics will be discussed in class. Tentative topics should be discussed with me during office hours in early February. Final topics are to be handed in by the end of February. Three sessions at the end of the semester will be devoted to **group presentations** of research papers.

Problems with computing: If you want to see me about a computing problem, bring a copy of the commands and output, along with a disk (preferably a flash/thumb drive). If you send a question electronically, include the commands and output. It is often impossible to diagnose error messages without these.

Policy on late submissions

Late submissions of assignments and the term paper will incur a penalty of 20% of the assignment's grade. Each additional 24-hour delay (including over the week-end) will incur an extra 20%. Since many assignments require the use of the computer, be warned that the probability of a computer being down is inversely related to the number of days until the assignment is due. Please send any late submissions through MyCourses e-mail as soon as possible.

“Students’ rights and responsibilities”

Attendance and participation in class discussions.

You are responsible for all announcements made in class and on MyCourses, as well as for all course materials given out in class. You should also check for new announcements or material on MyCourses at least weekly.

Policy Concerning the Rights of Students with Disabilities

If you have a disability please contact the instructor to arrange a time to discuss your situation. It would be helpful if you contact the Office for Students with Disabilities at 398-6009 before you do this.

Remise des travaux en français

Conformément à la Charte des droits de l’étudiant de l’Université McGill, chaque étudiant a le droit de soumettre en français ou en anglais tout travail écrit devant être noté (sauf dans le cas des cours dont l’un des objets est la maîtrise d’une langue).

Les étudiants de ce cours peuvent rédiger tous leurs travaux (incluant les examens) en français, mais doivent pour ce faire obtenir la **permission préalable** de la professeure. **Aucune permission rétroactive ne sera accordée.**

Policy for the Accommodation of Religious Holy Days

1. Students will not be penalized if they cannot write examinations or be otherwise evaluated on their religious holy days where such activities conflict with their religious observances.
2. Students who because of religious commitment cannot meet academic obligations, other than final examinations, on certain holy days are **responsible for informing their instructor, with two weeks’ notice of each conflict.**
3. When the requested accommodation concerns a **final examination, students are responsible for advising their faculty office as soon as possible and not later than the deadline for reporting conflicts.** Additional documentation confirming their religious affiliation may be requested.

Statement on academic integrity at McGill

“McGill University values academic integrity. Therefore all students must understand the meaning and consequences of cheating, plagiarism and other academic offences under the Code of Student Conduct and Disciplinary Procedures (see www.mcgill.ca/integrity for more information).”

“L’université McGill attache une haute importance à l’honnêteté académique. Il incombe par conséquent à tous les étudiants de comprendre ce que l’on entend par tricherie, plagiat et autres infractions académiques, ainsi que les conséquences que peuvent avoir de telles actions, selon le Code de conduite de l’étudiant et des procédures disciplinaires (pour de plus amples renseignements, veuillez consulter le site www.mcgill.ca/integrity).”

© Instructor generated course materials (e.g., handouts, notes, summaries, exam questions, etc.) are protected by law and may not be copied or distributed in any form or in any medium without explicit permission of the instructor. Note that infringements of copyright can be subject to follow up by the University under the Code of Student Conduct and Disciplinary Procedures.

"In the event of extraordinary circumstances beyond the University's control, the content and/or evaluation scheme in this course is subject to change."

SCHEDULE

Course overview, expectations, introduction to Stata

1. *MONDAY, JANUARY 6TH*

Descriptives and transformations

2. *WEDNESDAY, JANUARY 8TH*

Basic descriptive statistics and graphs. Skewness and nonlinearity. Transformations: rationale, family of powers and roots, Tukey's ladder, "bulging rule".

3. *MONDAY, JANUARY 13TH: Lab session*

Regression: OLS

4. *WEDNESDAY, JANUARY 15TH*

5. *MONDAY, JANUARY 20TH*

Types of variables and when the standard regression model may be inappropriate. Fitting a straight line. OLS basics. Statistical inference for simple regression.

6. *WEDNESDAY, JANUARY 22ND: Lab session*

7. *MONDAY, JANUARY 27TH*

Multiple regression/Multivariate OLS.

8. *WEDNESDAY, JANUARY 29TH: Lab session*

9. *MONDAY, FEBRUARY 3RD*

Categorical independent variables. Interaction terms. Relative importance.

10. *WEDNESDAY, FEBRUARY 5TH: Lab session*

11. *MONDAY, FEBRUARY 10TH*

OLS assumptions; BLUE. Who created OLS?

12. *WEDNESDAY, FEBRUARY 12TH: Lab session*

Assignment 1: OLS (February 12th). Due by e-mail (through MyCourses) by midnight.

Regression: binary outcomes

13. *MONDAY, FEBRUARY 17TH*

Linear probability model. Latent variable model for binary outcomes. Identification. Nonlinear probability model.

14. *WEDNESDAY, FEBRUARY 19TH: Lab session*

15. *MONDAY, FEBRUARY 24TH*

Cumulative Distribution Function. How to get probabilities. Graphs.

Final topics for team research papers due February 26th

16. *WEDNESDAY, FEBRUARY 26TH: Lab session*

Regression: ordinal outcomes

17. *MONDAY, MARCH 10TH*

Ways of dealing with ordinal outcomes: integer scores; summary scores/factor analysis/PCA; baseline category logit; adjacent logit; cumulative logit. Advantages of ordinal logit model over these alternatives. Latent variable interpretation. Estimation, interpretation.

Assignment 2: logistic regression (March 12th). Due by e-mail (through MyCourses) by midnight.

18. *WEDNESDAY, MARCH 12TH: Lab session*

19. MONDAY, MARCH 17TH

The parallel regression assumption. Generalized ordinal logit. Related models for ordinal outcomes.

Regression: nominal outcomes

20. WEDNESDAY, MARCH 19TH: Lab session

21. MONDAY, MARCH 24TH

The multinomial logit model. Computing and testing other contrasts. RRRs, interpretation. The IIA assumption.

22. WEDNESDAY, MARCH 26TH: Lab session

23. MONDAY, MARCH 31ST

Multinomial logit contd. Testing IIA assumption. Predicted probabilities, graphs.

Assignment 3: ordinal and multinomial logit regression (Monday, April 7th). Due by e-mail (through MyCourses) by midnight.

GROUP PRESENTATIONS OF RESEARCH PAPERS:

WEDNESDAY, APRIL 2ND

MONDAY, APRIL 7TH

WEDNESDAY, APRIL 9TH

TEAM RESEARCH PAPER (April 9TH).