I. COURSE DOMAIN AND BOUNDARIES

Generalized linear models (GLMs) are a collection of statistical methods used to analyze categorical and limited dependent variables. In this course, students will learn fundamental concepts and skills to conduct generalized linear models, and know how to apply these techniques to social, behavioral, and health research. The course covers the following topics: the Nelder and Wedderburn (1972) framework of generalized linear models and the key concept of “link function,” maximum likelihood estimator, logistic and probit models for binary outcomes, ordered logit and ordered probit models for ordinal outcomes, multinomial logit model for nominal outcomes, Poisson regression and negative binomial regression for count outcomes, other models for count outcomes, GEE and random effects model for clustered categorical outcomes, log-linear models, quasi-likelihood methods, and model fit/validation. Students taking this course are assumed to have taken statistics courses on inferential statistics and regression analysis, and know how to run Stata software package to perform statistical analysis.

II. COURSE OBJECTIVES

At the completion of the course, students will be able to:

1. Understand challenges posted by categorical and limited dependent variables; Know how to choose an appropriate statistical method to analyze such data;
2. Understand the statistical principles and know how to use Stata to analyze a binary nominal dependent variable by using the logistic regression model;
3. Understand the statistical principles and know how to use Stata to analyze a binary nominal dependent variable by using the probit model;
4. Understand the statistical principles and know how to use Stata to analyze an ordinal dependent variable by using ordered logit and ordered probit models;
5. Understand the statistical principles and know how to use Stata to analyze a nominal dependent variable by using the multinomial logit model;
6. Understand the statistical principles and know how to use Stata to analyze count dependent variable by using the Poisson regression model;
7. Understand the statistical principles and know how to use Stata to analyze count dependent variable by using the negative binomial regression model;
8. Understand the statistical principles and know how to use Stata to analyze count dependent variable by using other count models;
9. Understand the basic setup and statistical principles of Tobit and log-linear models;
10. Understand the problem presented by clustered categorical data and basic methods to address it;
11. Know how to read, evaluate, and criticize studies using categorical and limited dependent variables.

III. EDUCATIONAL OUTCOMES

By taking the course, the students should demonstrate:

1. The ability to employ at least five GLM methods to analyze a categorical dependent variable;
2. The ability to critically review studies using multivariate models that involve categorical data;
3. The ability to present study findings using GLM in an organized, succinct, and accurate fashion;
4. The readiness for taking advanced statistical methods.

IV. ORGANIZATION OF COURSE

The content of this course is covered through reading, lectures, writing, and working with the computer. Best results are obtained when the reading is completed prior to each class meeting, and homework assignments are completed by the due times. The course will teach the application of Stata software package to conduct GLM analyses.

V. ROLES OF INSTRUCTORS AND TEACHING ASSISTANT

Instructor: Class meetings will be primarily lecture/discussion format. The instructor will be responsible for the preparation and response to student inquiries regarding the readings, exercises, and the course project. Additionally, the instructor will be available during office hours and by special appointment for questions regarding the course.
Teaching Assistant: The TA will lead most computer lab sessions to help students practice the software programs pertaining to topics covered by weekly lectures. The TA will grade students’ assignments and term papers based on guidelines provided by the instructor. The TA will hold weekly office hours.

Students should consult with the instructor for any disputable issue regarding grading. The instructor may or may not change the grade based on his evaluation. In this situation, the instructor’s evaluation and grade will be final.

VI. ROLES OF STUDENTS

This is an applied course that requires significant attention to out-of-class activities as well as readings. Students will get lost without regular attention to readings and attendance at all scheduled meetings. Students are expected to seek assistance and clarification when needed, complete assigned readings and exercises, and provide feedback regarding the effectiveness of the class.

Students are encouraged to participate in group discussions through learning communities. However, all assignments and term papers should be completed independently. Academic integrity in the completion of oral presentations and written assignments is expected. Violations of academic integrity (e.g., plagiarism) will result in notification to the Associate Dean for the PhD program and advisor.

VII. WASHINGTON UNIVERSITY ACADEMIC SUPPORT POLICIES

Accommodations based upon sexual assault: The University is committed to offering reasonable academic accommodations to students who are victims of sexual assault. Students are eligible for accommodation regardless of whether they seek criminal or disciplinary action. Depending on the specific nature of the allegation, such measures may include but are not limited to: implementation of a no-contact order, course/classroom assignment changes, and other academic support services and accommodations. If you need to request such accommodations, please direct your request to Kim Webb (kim_webb@wustl.edu), Director of the Relationship and Sexual Violence Prevention Center. Ms. Webb is a confidential resource; however, requests for accommodations will be shared with the appropriate University administration and faculty. The University will maintain as confidential any accommodations or protective measures provided to an individual student so long as it does not impair the ability to provide such measures.

If a student comes to me to discuss or disclose an instance of sexual assault, sex discrimination, sexual harassment, dating violence, domestic violence or stalking, or if I otherwise observe or become aware of such an allegation, I will keep the information as private as I can, but as a faculty member of Washington University, I am required to immediately report it to my Department Chair or Dean or directly to Ms. Jessica Kennedy, the University’s Title IX Coordinator. If you would like to speak with the Title IX Coordinator directly, Ms. Kennedy can be reached at (314) 935-3118, jwkennedy@wustl.edu, or by visiting her office in the Women’s Building. Additionally, you can report incidents or complaints to Tamara King, Associate Dean for Students and Director of Student Conduct, or by contacting WUPD at (314) 935-5555 or your local law enforcement agency. You can also speak confidentially and learn more about available
resources at the Relationship and Sexual Violence Prevention Center by calling (314) 935-8761 or visiting the 4th floor of Seigle Hall.

**Bias Reporting:** The University has a process through which students, faculty, staff and community members who have experienced or witnessed incidents of bias, prejudice or discrimination against a student can report their experiences to the University’s Bias Report and Support System (BRSS) team. See: brss.wustl.edu

**Mental Health:** Mental Health Services’ professional staff members work with students to resolve personal and interpersonal difficulties, many of which can affect the academic experience. These include conflicts with or worry about friends or family, concerns about eating or drinking patterns, and feelings of anxiety and depression. See: shs.wustl.edu/MentalHealth.

**VIII. GRADES**

Students are required to complete five assignments and a final exam. Weighting for the final course grade is shown below:

<table>
<thead>
<tr>
<th>Component</th>
<th>Grade Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assignment 1</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 2</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 3</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 4</td>
<td>12%</td>
</tr>
<tr>
<td>Assignment 5</td>
<td>20%</td>
</tr>
<tr>
<td>Final Exam (take home)</td>
<td>32%</td>
</tr>
</tbody>
</table>

**Grading scale:** the overall grade will be calculated according to the following scale:

- 98-100 A+ 77-79 C+ <60 F
- 95-97 A 74-76 C
- 90-94 A- 70-73 C-
- 87-89 B+ 67-69 D+
- 84-86 B 64-66 D
- 80-83 B- 60-63 D-

**IX. TEXTBOOKS AND READINGS**

**Required Textbooks**


**Required Articles for Reading**

Required journal articles will be available on Blackboard and University’s online library.
X. CLASS SCHEDULE AND READING ASSIGNMENTS

8-28-19  Session 1: Introduction and Course Overview
1. Why categorical data analysis?
2. Types of measurement and definition of categorical variables
3. History of the development of categorical analysis
4. Review of the linear regression model
5. Course overview

Readings:
Long, Chapters 1 & 2.
Long & Freese, Chapter 1

9-4-19  Session 2: Preliminaries
1. Exponential and logarithmic functions
2. Contingency tables
3. Relative risk and odds ratio
4. Chi-squared distribution
5. Chi-square test
6. Fisher’s exact test

Readings:
Course PPT and lab documents

Assignment 1 out (Due: 9/18/19): (1) exercises on solving problems of exponential and logarithmic functions; (2) exercises on performing chi-square test and Fisher’s exact test; and (3) design a research project that requires GLMs.

9-11-19  Session 3: Fundamental Stata Commands for Analyzing GLMs
(This session will be conducted by Rachel Garg)
1. Stata basics
2. Stata commands for estimating, testing, and fit of GLMs
3. Using factor-variable notation (i.e., prefixing a variable name with i)

Readings:
Long & Freese, Chapters 2, 3, & 4
(Note: you are required to read all three chapters. Yi will cover the basics of Chapters 2 & 3 in her class lecture. You don’t have to understand every command of Chapter 4. Shenyang will discuss and give examples of key commands of Chapter 4 in future class sessions.)

9-18-19  Session 4: Generalized Linear Models & Maximum Likelihood Estimator
1. The Nelder and Wedderburn (1972) framework of generalized linear models
2. The concept of “link function”
3. Review of binary logistic regression
4. The maximum likelihood estimator

Readings:
Long, Sections 3.5 & 3.6 in Chapter 3

Assignment 2 out (Due: 10/9/19): (1) exercises on solving problems about linear probability, probit, and logit models; (2) use provided data to run probit and logistic regression models.

9-25-19 Session 5: Binary Outcomes: The Linear Probability and Probit Models
1. The linear probability model
2. The probit model
3. Running the linear probability and probit models with Stata

Readings:
Long, Chapter 3
Long & Freese, Chapter 5

10-2-19 Session 6: Binary Outcomes: the Logistic Regression and Interpretation of Results of Binary-Outcome Models
1. Summary of the important issues running the binary logistic regression
2. Marginal effect at the mean, marginal effect at representative values, and average marginal effects
3. The graphic methods

Readings:
Long, Chapter 4
Long & Freese, Chapter 6
10-9-19  Session 7: Ordinal Outcomes: the Ordered Logit and Ordered Probit Models
1. A latent-variable model
2. Ordered logit and ordered probit models
3. Estimation and hypothesis testing
4. Interpretations using predicted probabilities

Readings:
Long, Chapter 5
Long & Freese, Chapter 7

Assignment 3 out (Due: 10/30/19): (1) exercises on solving problems about ordered and multinomial logit models; (2) use provided data to run ordered and multinomial logit models.

10-16-19  Session 8: Nominal Outcomes: the Multinomial Logit and Related Models
1. The multinomial logit model
2. Estimation and hypothesis testing
3. Interpretations
4. The graphic methods
5. Related models

Readings:
Long, Chapter 6
Long & Freese, Chapter 8
Session 9: Count Outcomes: the Poisson Regression Model
1. Tests of skewed data
2. The Poisson distribution
3. The Poisson regression model
4. Estimation and hypothesis testing
5. Interpretations
6. The graphic methods

Readings:
Long, Sections 8.1 & 8.2 in Chapter 8
Long & Freese, Sections 9.1 & 9.2 in Chapter 9

Session 10: Count Outcomes: the Negative Binomial Regression Model
1. The negative binomial regression model
2. Estimation and hypothesis testing
3. Interpretations
4. The graphic methods

Readings:
Long, Section 8.3 in Chapter 8
Long & Freese, Section 9.3 in Chapter 9

Assignment 4 out (Due: 11/20/19): (1) multiple-choices questions; (2) use provided data to run Poisson and negative binomial regression models.

Session 11: Review -- The Essentials of Binary Probit & Logit, Ordered Logit, Multinomial Logit, Poisson Regression, and Negative Binomial Regression Models
1. The basic setup of each model
2. Description of analytic methods
3. Interpretation using predicted probabilities (MEMs, MERs, AMEs)
4. Interpretation using odds ratios and incident-rate ratios
5. Hypothesis testing
6. Presentation of tables and graphs

Readings:
Class PPTs and Lab notes for Sessions 6, 7, 8, 9, & 10

Assignment 5 out (Due: 12/4/19): Students are organized into several groups. Each group is provided with two article using GLMs. You are required to convene two group-discussion sessions to evaluate the strengths and limitations of the provided study. Present your findings to the class on the due date.

Final Exam out (Due: 12/15/19): Use a dataset provided by the course or dataset you choose to run analysis with a GLM and a competing model. Depending on the nature of your outcome variable, you should choose one of the following three pairs of models: binary logistic regression and probit model, ordered logistic regression and multinomial logit model, or Poisson regression and negative binomial regression. Write a paper (no more than 20 pages, double spaced) to present findings. The paper should include: (1) description of research questions and the study data; (2) specifications of the main statistical model and the competing model; (3) presentation of descriptive statistics of the sample data; (4) tests of important hypotheses including hypotheses about interaction terms; (5) presentation of the main models using graphic methods; (6) interpretation of main findings; and (7) discussion.

11-13-19 Session 12: Topics to Be Determined, A Session Arranged to Adjust Course Pace

11-20-19 Session 13: Advanced Topics – Other Categorical Models

1. The Tobit model for censored outcomes
2. Log-linear models
3. The problem presented by clustered data
4. The marginal approach: GEE
5. The random effect models

Readings:
Long, Chapters 7 and 9
11-27-19  Thanksgiving break, no class

12-4-19  Session 14: Student Presentations and Course Conclusions

1. Student presentations
2. Course conclusions

12-15-19  Final Exam Due