PLS415/515: Maximum Likelihood Models

Fall 2020

Gento Kato
gento.kato@nu.edu.kz

Nazarbayev University

Last Update: August 16, 2020
(**TENTATIVE and SUBJECT TO CHANGE**)

1 General Information

Class Hours

Official Class Hours for this class is **1:30-4:20pm Tuesdays (Nur-Sultan local time)**. However, given that the course is remotely taught, following adjustments are planned to be made (subject to change as the semester proceeds):

- **Theoretical contents** of the course will be offered in a format of **a video presentation**. In each week (except for the first week), a video is planned be uploaded before 1:30pm of Tuesday (Nur-Sultan Local Time).
- **Applied contents (i.e., implementation of methods in R)** will be offered in a format of **video conference**, during **2:30-3:30pm Tuesdays (Nur-Sultan local time)**. Students are expected to attend, but in case one has technical troubles joining a live conference, a video recording will be uploaded after the class.
- During **3:30-4:20pm Tuesdays**, a video conference will remain open as long as students have questions to ask about the content of class.

Office Hours

Office hours will be held over Zoom and it is **by appointment only**. Schedule a meeting by checking the availability from the URL below.

https://www.meetingbird.com/l/gentok/office-hours

Each meeting slot is for 15 minutes. If you want longer meeting, reserve two slots separately. The Zoom invitation will be sent to your E-mail address once you successfully completed the reservation. **If the given slots do not work for you, contact me directly with the proposal of your available times**.

2 Prerequisite

It is assumed that students have learned the materials covered in PLS210 & 211 or PLS 510 & 511. Specifically, students are assumed to have basic knowledge on research design, probability and statistical hypothesis testing. Prior knowledge of linear regression (i.e., ordinary least square: OLS) is also assumed, while we review it in class.

3 Course Objectives

In many quantitative research in the field political science, a variable of interest is discrete, truncated, or non-normally distributed. It is difficult to apply linear model and ordinary least squares (OLS) regression to such data, because assumptions in OLS may be violated. This course introduces a more general framework called "generalized linear models" (GLM), which is flexible and applicable to wider range of data. Maximum likelihood is a technique that is used to estimate such models.

After completing this course, students are expected to acquire the skills to ...

1. Understand the mechanics of GLM and Maximum Likelihood.
2. Interpret analytical results of GLM.
3. Apply GLM to real-world data.

The course especially focuses on applied side of GLM and Maximum likelihood. Skills acquired from this course help to develop research idea and modeling strategy for various types of political science data.

4 Textbook

This main textbooks for this course are:

- J. Scott Long and Simon Cheng. 2004. "Regression Models for Categorical Outcomes." In *Handbook of Data Analysis,* edited by Melissa A. Hardy and Alan Bryman, 259–284. SAGE Publications, Ltd
- Alan Agresti. 2015. *Foundations of Linear and Generalized Linear Models.* Somerset: Wiley. https://ezproxy.nu.edu.kz:2077/lib/astana/detail.action?docID=1895564
- Janet M. Box-Steffensmeier and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists.* New York, NY: Cambridge University Press. https://ezproxy.nu.edu.kz:2364/10.1017/CBO9780511790874

Long and Cheng (2004) is available through Moodle and Agresti (2015) and Box-Steffensmeier and Jones (2004) are available online through NU library (by clicking the URLs above). You will be prompted to enter your NU user ID and password to access materials. There will be additional required readings that will be available through Moodle.

5 Statistical Software

In this course we use the open-source statistical software R (http://www.r-project.org). On your own computer, you are required to install R . It is also recommended to install RStudio (http://www.rstudio.com/)—a user interface that simplifies many common operations. While R codes will be provided through Moodle, online resources such as R Cookbook (https://rc2e.com/index.html) would be a great resource for you to start out using R and RStudio.

6 Course Requirements

There are four components to the course requirements:

1. **R Programming Assignments (10% * 5 = 50%)**: There will be (almost) bi-weekly programming assignments.
2. **Replication Assignment (15%)**: There will be a replication assignment due on **Friday, October 2** in **Week 7**. In the assignment, students will be asked to replicate a result of existing political science study. Additional details about the format will be provided by **Week 4**.
3. **Research Project (30%)**: There will be a original research project assignment due on **Monday, November 30**. Students are expected to test original hypothesis using one of the methods covered in the class. Additional details about the format will be provided as the semester proceeds.
4. **Attending Consultation Meeting for Research Project (5%)**: Students are expected to consult with me regarding the plan for research project through video-meeting. The attendance in this meeting will count as 5% of grade.

7 Course Outline

Schedule and contents are subject to change.

Introduction

Week 1: August 18

- Reading: Agresti (2015) Ch.1
- R Assignment 1: Due Monday, August 31.

Review of Linear Regression

Week 2: August 25

- Reading: Agresti (2015) Ch.2

- Recommended:
  - Ross M. Stolzenberg. 2004. "Multiple Regression Analysis." In *Handbook of Data Analysis,* 165–207. Available through Moodle. SAGE Publications, Ltd

Models for Binary Data: Logit/Probit

Week 3: September 1

- Reading:
  - Long and Cheng (2004) pp.2-16
  - Agresti (2015) Ch.4-5
- R Assignment 2: Due Monday, September 14.

Week 4: September 8

Models for Ordinal Data

Week 5: September 15

- Reading:
  - Long and Cheng (2004) pp.17-31
  - Agresti (2015) Ch.6.2, 6.3.3
- R Assignment 2: Due Monday, October 12.

Models for Nominal Data

Week 6: September 22

- Reading:
  - Long and Cheng (2004) p.31-40
  - Agresti (2015) Ch.6.1, 6.3.1, 6.3.2

Week 7: September 29

- Replication Assignment: Due Friday, October 2.

FALL BREAK

Models for Count Data

Week 8: October 13

- Reading:
  - Agresti (2015) Ch.7
  - C. E. Rose et al. 2006. "On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data." *Journal of Biopharmaceutical Statistics* 16, no. 4 (August 1): 463–481. Accessed April 14, 2020. doi:10.1080/10543400600719384. pmid: 16892908. https://doi.org/10.1080/10543400600719384
- R Assignment 3: Due Monday, October 26.

Censoring, Truncation and Sample Selection Regression Models

Week 9: October 20

- Reading:
  - G. S Maddala. 1983a. "Chapter 6: Censored and Truncated Regression Models." In *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press
  - James J. Heckman. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161. doi:10.2307/1912352
  - Jeffery A. Dubin and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods & Research* 18, nos. 2-3 (November 1): 360–390. doi:10.1177/0049124189018002006
- R Assignment 3: Due Monday, October 26.

Event History Models

Week 10: October 27

- Reading: Box-Steffensmeier and Jones (2004) Ch.2-6, 8
- R Assignment 4: Due Monday, November 9.

Week 11: November 3

Visual Presentation of Results

Week 12: November 10

- R Assignment 5: Due Monday, November 23.

Week 13: November 17

Week 14: November 24

- Research Project: Due Monday, November 30.

8 Grading Policy

8.1 Grading Scale

Each assignments will be given a **letter grade**, either by A to F or Check plus to minus scale. Each letter grade is translated to grade point by the following tables:

A to F Scale

Letter Grade	Grade Point
A	4.00
A−	3.67
B+	3.33
B	3.00
B−	2.67
C+	2.33
C	2.00
C−	1.67
D+	1.33
D	1.00
F	0.00

Check Plus to Minus Scale

Letter Grade	Grade Point
✓+	4.00
✓	3.00
✓−	2.00
Fail	0.00

The final letter grade in A to F scale will be determined by the weighted average of grade points according to the percentages presented in Course Requirements section.

8.2 Late Submission

For any assignment, the following late submission policies will be applied. The assignment submitted incorrectly will be considered as missing.

- Late submission **within 24 hours of the deadline**: Allowed without penalty. However, **a student needs to notify me by E-mail before the deadline**. Without notification, the assignment will be treated in the same way as the *late submission within 1 week of the deadline*.
- Late submission **within 1 week of the deadline**: Allowed with penalty. **The highest grade a student can get will be B or ✓.** No notification is required.
- Late submission **after 1 week of the deadline**: Not allowed at all time.

If you have any special reasons that force you to submit after the deadline, please E-mail me or come talk to me **before the deadline**. If the reason is valid, I may extend the deadline for you. I will **not accept any request after the deadline**.

8.3 Regrading Request

If there is a very clear error in grading your assignment or exam, please let me know as soon as possible. However, if you wish to contest your grade on other grounds, then you must submit a written request to me with following contents:

- Clearly identify the components of the exam or assignment that should be reconsidered.
- For each component, **propose the score** that you thing you deserve on that specific component.
- For each component, provide at least a paragraph length explanation for why you think your score should be reconsidered.

I will only have a right to **accept or reject** your proposal. If your proposed score and explanation are compelling, I will change your score on that specific component to the score you proposed. Otherwise, I will reject it and keep the score as it is.

Under no circumstances will I change any score on any assignment ***more than one week*** *after we have released the grades to the class. If you anticipate that it will take you more than a week to review your graded assignment and draft a response, then you will need to request an extension in advance.*

9 Academic Conduct

I do not expect that any of the students in this course will violate the University's Student Code of Conduct. However, please be aware that **cheating, plagiarism, or other violations of the Code will not be tolerated and will be reported to the Vice Dean and/or the Dean** for appropriate action. Please check NU Student Code of Conduct for more details (https://nu.edu.kz/wp-content/uploads/2017/11/NU-Student-Code-of-Conduct.pdf). Also, NU library provides a helpful guideline regarding the academic integrity (https://nu.kz.libguides.com/avoiding_plagiarism/academicintegrity).

10 A Safe and Comfortable Learning Environment

As an instructor, one of my responsibilities is to help create a safe and comfortable learning environment for my students and for the university as a whole. If you have any problems or concerns regarding your student life, including, but not limited to relationships, stress, self-esteem, body image, sexuality, anxiety, and/or depression, Psychological Counselling Center (PCC) provides consultation service for you. Contact nu_counseling@nu.edu.kz or make an appointment with psychologist through NU web portal, under `SERVICES` → `Consultation` tab. PCC also has YouTube Channel (https://www.youtube.com/channel/UCZZCN6Kc7UcwWJO4Ai3yOXQ) and Instagram (https://www.instagram.com/nu_pcc/) with additional resources.

11 Alternative Textbooks

Following short list is for you get some alternative perspectives on the topic. They are *not* required, and *may not* be available online.

Alan Agresti. 2019. *An Introduction to Categorical Data Analysis.* 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.

John P Hoffmann. 2016. *Regression Models for Categorical, Count, and Related Variables: An Applied Approach.* University of California Press

Tamás Rudas. 2018. *Lectures on Categorical Data Analysis.* Springer

Russell B Millar. 2011. *Maximum Likelihood Estimation and Inference with Examples in R, SAS, and ADMB.* Chichester, Sussex, U.K.: Wiley. https://ezproxy.nu.edu.kz:2077/lib/astana/reader.action?docID=697473&ppg=1

Gary King. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* University of Michigan Press

Scott R Eliason. 1993. *Maximum Likelihood Estimation: Logic and Practice.* Newbury Park, Calif.; London: Sage

G. S Maddala. 1983b. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press. https://ezproxy.nu.edu.kz:2364/10.1017/CBO9780511810176

References

Agresti, Alan. 2015. *Foundations of Linear and Generalized Linear Models.* Somerset: Wiley. https://ezproxy.nu.edu.kz:2077/lib/astana/detail.action?docID=1895564.

———. 2019. *An Introduction to Categorical Data Analysis.* 3rd ed. Hoboken, NJ: John Wiley & Sons, Inc.

Box-Steffensmeier, Janet M., and Bradford S. Jones. 2004. *Event History Modeling: A Guide for Social Scientists.* New York, NY: Cambridge University Press. https://ezproxy.nu.edu.kz:2364/10.1017/CBO9780511790874.

Dubin, Jeffery A., and Douglas Rivers. 1989. "Selection Bias in Linear Regression, Logit and Probit Models." *Sociological Methods & Research* 18, nos. 2-3 (November 1): 360–390. doi:10.1177/0049124189018002006.

Eliason, Scott R. 1993. *Maximum Likelihood Estimation: Logic and Practice.* Newbury Park, Calif.; London: Sage.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161. doi:10.2307/1912352.

Hoffmann, John P. 2016. *Regression Models for Categorical, Count, and Related Variables: An Applied Approach.* University of California Press.

King, Gary. 1998. *Unifying Political Methodology: The Likelihood Theory of Statistical Inference.* University of Michigan Press.

Long, J. Scott, and Simon Cheng. 2004. "Regression Models for Categorical Outcomes." In *Handbook of Data Analysis,* edited by Melissa A. Hardy and Alan Bryman, 259–284. SAGE Publications, Ltd.

Maddala, G. S. 1983a. "Chapter 6: Censored and Truncated Regression Models." In *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press.

———. 1983b. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge: Cambridge University Press. https://ezproxy.nu.edu.kz:2364/10.1017/CBO9780511810176.

Millar, Russell B. 2011. *Maximum Likelihood Estimation and Inference with Examples in R, SAS, and ADMB.* Chichester, Sussex, U.K.: Wiley. https://ezproxy.nu.edu.kz:2077/lib/astana/reader.action?docID=697473&ppg=1.

Rose, C. E., S. W. Martin, K. A. Wannemuehler, and B. D. Plikaytis. 2006. "On the Use of Zero-Inflated and Hurdle Models for Modeling Vaccine Adverse Event Count Data." *Journal of Biopharmaceutical Statistics* 16, no. 4 (August 1): 463–481. Accessed April 14, 2020. doi:10.1080/10543400600719384. pmid: 16892908. https://doi.org/10.1080/10543400600719384.

Rudas, Tamás. 2018. *Lectures on Categorical Data Analysis.* Springer.

Stolzenberg, Ross M. 2004. "Multiple Regression Analysis." In *Handbook of Data Analysis,* 165–207. Available through Moodle. SAGE Publications, Ltd.