Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

Applied Regression II
P8110

CLASS SESSIONS
Thursdays, 8:30 – 11:20AM, Hammer 312

INSTRUCTOR
Qixuan Chen, PhD
Assistant Professor
Email: qc2138@cumc.columbia.edu
Office: 722 West 168th Street, R644

TEACHING ASSISTANTS
Shar Barracks (szw2101@cumc.columbia.edu)
Office hours & location: Wednesdays 12:00-1:00PM at MSPH R638

Peng Wu (pw2394@cumc.columbia.edu)
Chubing Zeng (cz2337@cumc.columbia.edu)
Office hours & location: by appointments

COURSE DESCRIPTION

This course will introduce the statistical methods for analyzing censored data, non-normally distributed response data, and repeated measurements data that are commonly encountered in medical and public health research. Topics include estimation and comparison of survival curves, Cox models for survival data, multinomial logistic regression models for nominal outcomes, cumulative logistic regression models for ordinal outcomes, Poisson and negative binomial models for count data, generalized estimating equations, and mixed models. Examples are drawn from the health sciences.

PREREQUISITES

This course is designed for students who have already completed P8100 (Applied regression analysis I) and have some programming experiences of statistical software package SAS.

COURSE LEARNING OBJECTIVES

Students who successfully complete this course will be able to:

- Identify types of censoring in survival data analysis, create Kaplan-Meier survival curves, generate descriptive statistics for survival data, compare survival functions among two or more groups, and use Cox models to analyze survival data.
- Use logistic regression models to analyze binary, nominal, or ordinal response data, and use Poisson and negative binomial regression models to analyze count response data.
- Use generalized estimating equations and mixed models to analyze repeated measures data.
- Write simple SAS programs and read SAS outputs.
- Write statistical analysis section, create tables and figures, and interpret statistical analysis results for scientific papers.

TEXTBOOKS

Required textbook: None

Recommended textbooks:

- Allison PD (2010). *Survival Analysis Using SAS: A Practical Guide Second Edition*. SAS Institute. [eBook]
- Hosmer DW, Lemeshow S and May S (2008). *Applied Survival Analysis: Regression Modeling of Time-to-Event Data*. John Wiley & Sons, Inc.: New York. [eBook]
- Dupont WD (2009). *Statistical Modeling for Biomedical Researchers: A Simple Introduction to the Analysis of Complex Data*. Cambridge University Press. [eBook]
- Dobson AJ (2001). *An Introduction to Generalized Linear Models, 2nd Edition*. Chapman & Hall. [eBook]

Recommended lecture notes:

- Davidian M. Applied Longitudinal Data Analysis, Lecture Notes. http://www.stat.ncsu.edu/people/davidian/courses/st732/notes/toc.pdf

ASSESSMENT AND GRADING POLICY

The course will consist of lectures, homework assignments, quizzes, and two projects. Grading will be based on: homework (30%), quiz (30%), midterm project (20%), and final project (20%).

- The homework should be submitted via the Columbia CourseWorks. It is possible to submit homework up to two days late, but there is a 10% penalty for each day. Students are encouraged to work together on homework, although only in the spirit of learning. There should be no copying of work.
- The quizzes will be closed book closed notes, but allow bringing a one-page cheat sheet.
- Students will be divided into groups with 3 students per group. Students in each group will work together on a midterm project (survival data analysis) and a final project (non-normal data or repeated measures data analysis), and submit a written report for each project. Each group will be randomly assigned to do an in-class presentation on either their midterm project or final project.

Final letter grades will be assigned on the basis of traditional scoring: top 5% of the class A+; ≥ 90% A; 85-90% A-; 80-85% B+; 75-80% B; 70-75% B-; 65-70% C+; 60-65% C; and < 60% F.

ADDITIONAL COURSE INFORMATION

We will use SAS. SAS is available for use on PCs in the library and in the Student Learning Center (LC 17-107). SAS is available for student purchase by going to http://www.cubhis.org/getting_help/sas.html.

All course notes and handouts will be posted on the COURSEWORKS web site. It is students' responsibility to print out all materials before coming to class. Hard copies of handouts will NOT be provided in class. All students are required to arrive on time. Cell phones must be turned off, set to vibrate/silent. No phone calls are allowed in the class.

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Academic Integrity

Students are required to adhere to the Mailman School Honor Code, available online at http://mailman-handbook.com/2009/node/165

Disability Access

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). Students who have, or think they may have a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu. If you have already registered with ODS, please speak to your instructor to ensure that she has been notified of your recommended accommodations by Lillian Morales (lm31@columbia.edu), the School's liaison to the Office of Disability Services.

COURSE SCHEDULE

Week 1 – Syllabus & Estimating Survival Functions	
Jan 21	Learning Objectives: ▪ Syllabus ▪ Definition of time-to-event and censoring ▪ Survival function estimator without censoring ▪ Survival functions with censoring using Kaplan-Meier method ▪ Kaplan-Meier survival curves ▪ Greenwood's variance estimator for survival functions ▪ Confidence interval estimates for survival functions

Week 2 – Quantile and Mean Estimation, Comparison of Survival Functions	
Jan 28	Assignment due: HW # 1 Learning Objectives: ▪ Review HW # 1 ▪ Quantile estimation ▪ Confidence intervals for quantiles ▪ Mean survival time estimation ▪ SAS syntax and outputs for survival functions, quantiles, and means ▪ Comparison of survival functions using log-rank and Wilcoxon test

Week 3 – Cox Models	
Feb 4	Assignment due: HW # 2 Learning Objectives: ▪ Review HW # 2 ▪ SAS syntax and outputs for survival functions comparison ▪ Hazard functions ▪ Introduction to the PH models ▪ Cox model interpretation ▪ Nominal scale covariates and continuous scale covariates ▪ Confounders and interaction effects

Week 4 – Cox Models	
Feb 11	Quiz # 1 Learning Objectives: ▪ Cox Model comparison ▪ Analyze a real data using survival data analysis methods

Week 5 – Survival Data Analysis Case Study	
Feb 18	Assignment due: HW #3 Learning Objectives: ▪ Review HW # 3 ▪ Review Quiz # 1 ▪ Review midterm project ▪ Analyze a real data using survival data analysis methods ▪ Time-dependent covariate

Week 6 – PH Assumption Checking & Logistic Regression	
Feb 25	Assignment due: HW # 4 Learning Objectives: ▪ Review HW # 4 ▪ Model checking for the PH models ▪ Interactions with time ▪ Stratification ▪ Introduction to GLM ▪ Logistic regression model

Week 7 – Multinomial Logistic Regression	
Mar 3	Quiz # 2 Learning Objectives: ▪ Logistic regression model diagnostics ▪ Logistic regression model for nominal outcomes

Week 8 – Midterm project	
Mar 10	Report due: midterm project Learning Objectives: ▪ Midterm project presentation ▪ Review Quiz #2

Week 9 – NO CLASS	
Mar 17	Spring Break

Week 10 – Ordinal logistic regression and Poisson regression	
Mar 24	Learning Objectives: ▪ Ordinal logistic regression ▪ Poisson regression model

Week 11 – GEE	
Mar 31	Assignment due: HW # 5 Learning Objectives: ▪ Review HW # 5 ▪ Introduction to GEE ▪ SAS syntax and outputs for GEE models ▪ GEE case studies

Week 12 – Random Intercept Model	
Apr 7	Quiz # 3 Learning Objectives: ▪ Random intercept model ▪ Review final project

Week 13 – Random Slope Model and Linear Mixed Effects Model	
Apr 14	Assignment due: HW # 6 Learning Objectives: ▪ Review HW # 6 ▪ Review Quiz 3 ▪ linear mixed models

Week 14 – Linear Mixed Effects Model and Missing Data	
Apr 21	Quiz # 4 Learning Objectives: ▪ LMM: Case study ▪ Missing data

Week 15 – Final project presentation	
Apr 28	Report due: final project Learning Objectives: ▪ Review Quiz 4 ▪ Final project presentation