COLUMBIA UNIVERSITY | MAILMAN SCHOOL of PUBLIC HEALTH

BIOSTATISTICS

Biostatistical Methods II

P8131

COURSE DESCRIPTION

Regression analysis is widely used in biomedical research. Non-continuous (e.g., binary or count-valued) responses, correlated observations, and censored data are frequently encountered in regression analysis. This course will introduce advanced statistical methods to address these practical problems. Topics include generalized linear models (GLM) for non-Gaussian response, mixed-effects models and generalized estimating equations (GEE) for correlated observations, and Cox proportional hazards models for survival data analysis. Examples are drawn from real applications.

LEARNING OBJECTIVES

Students who successfully complete this course will be able to:

- Use GLM to analyze non-Gaussian response data;
- Use GEE and mixed models to analyze repeated measures;
- Understand the Kaplan-Meier estimator and use Cox models to analyze survival data;
- Use statistical software (e.g., R) to estimate model parameters, test hypotheses, and conduct model diagnostics;
- Summarize data analysis result and report findings.

INSTRUCTOR

Bin Cheng, Ph.D.
Professor of Biostatistics at CUMC
Phone: 212-342-1238
Email: bc2159@cumc.columbia.edu
Office: 722 West 168th Street, Room 648
Office hours: Fridays 3:00pm-4:00pm or by appointment

TEACHING ASSISTANTS

Zexi Cai: zc2626@cumc.columbia.edu; Office Hours:
Yuqi Miao: ym2771@cumc.columbia.edu; Office Hours:
Amy Pitts: ajp2257@cumc.columbia.edu; Office Hours:
Baoyi Shi: bs3141@cumc.columbia.edu; Office Hours:
Ziyi Wang: zw2716@cumc.columbia.edu; Office Hours:

CLASS SESSIONS

Mondays 5:30pm – 8:20pm, VEC 201

ZOOM INFORMATION

Link: https://columbiacuimc.zoom.us/j/91546770932
Meeting ID: 915 4677 0932
Passcode: P8131-001

EXAMS

Midterm exam: **February 28** (Monday), 5:30pm – 6:50pm (80 minutes);
Final exam: **May 2** (Monday), 5:30pm – 8:30pm (180 minutes).

PREREQUISITES

Students registering for this course **must** have completed Biostatistical Methods I, Probability and Inference I. Experience in **R** or other statistical software is **required**.

REFERENCES

Required textbooks:

Dobson A. J. and Barnett, A.G. (2008)

An Introduction to Generalized Linear Model. 3rd Ed. Chapman & Hall/CRC.

Fitzmaurice, G.M., Laird N.M. and Ware J.H. (2011)

Applied Longitudinal Analysis, 2nd Ed. Wiley.

Hosmer DW, Lemeshow S and May S (2008).

Applied Survival Analysis: Regression Modeling of Time-to-Event Data. 2nd Ed. Wiley.

Recommended textbooks:

McCullagh and Nelder (1989)

Generalized Linear Models. 2nd Ed. Chapman & Hall.

Faraway, J.J. (2016)

Extending the Linear Model with R, 2nd Ed. Chapman & Hall.

Diggle, P., Heagerty, P., Liang, K.Y., Zeger, S. (2013)

Analysis of Longitudinal Data, 2nd Ed. Oxford.

Klein, J.P. and Moeschberger M.L. (2003)

Survival Analysis: Techniques for Censored and Truncated Data, 2nd Ed. Springer

ASSESSMENT AND GRADING POLICY

Student grades will be based on the following:

Homework ---------- 40%
Mid-term exam ---------- 30%
Final exam ---------- 30%

There are ten weekly homework assignments, graded individually. No late homework accepted. Collaboration on homework assignments is allowed, but all submissions must be completed independently. There should be no copying of work.

The mid-term and final exams are CLOSE books and CLOSE notes. Students can bring **one** A4-sized cheat sheet to the mid-term, and **two** to the final (double sided, handwritten).

The final course grade will be determined using the School's letter grade system. Grades are A, B, C, with + and – as applicable. Grades are defined as follows:

A+ Reserved for highly exceptional achievement.
A Excellent. Outstanding achievement.
A- Excellent work, close to outstanding.
B+ Very good. Solid achievement expected of most graduate students.
B Good. Acceptable achievement.
B- Acceptable achievement, but below what is generally expected of graduate students.
C+ Fair achievement, above minimally acceptable level.
C Fair achievement, but only minimally acceptable.

C- Very low performance.
F Failure. Course usually may not be repeated unless it is a required course.

COURSE REQUIREMENTS

- Students are expected to attend all classes
- Late homework assignments are not acceptable unless a legible reason is given before the due day
- All homework should be turned in electronically on Canvas, unless otherwise noticed
- No incomplete grades will be given for this course

COURSE STRUCTURE: Lectures

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Students and faculty have a shared commitment to the School's mission, values and oath. http://mailman.columbia.edu/about-us/school-mission/

Academic Integrity

Students are required to adhere to the Mailman School Honor Code, available online at http://mailman.columbia.edu/honorcode.

Disability Access

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 s/he 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

Please see the lecture section of Courseworks to download the readings, exams, and lecture slides.

Lecture 1 – Review of Linear Regression

Learning Objectives:

- Review linear regression models, estimation, and diagnostics

Reading:

- Chapter 1, 2 of Dobson and Barnett

Homework: None

Lecture 2 – Introduction of Exponential Family Distributions

Learning Objectives:

- Review Binomial, Poisson, Exponential distributions
- Define general exponential family distributions

Reading:

- Chapter 3.1, 3.2, 3.3 of Dobson and Barnett

Homework: Assignment 1 (see below for details)

Lecture 3 – Introduction of GLM

Learning Objectives:

- Get familiar with the basic concepts of GLM (link functions, linear predictor, and random component)
- Review GLM examples

Reading:

- Chapter 3.4, (4.3 and 5) of Dobson and Barnett

Homework: Assignment 1

Lecture 4 – Introduction of GLM

Learning Objectives:

- Learn model fitting, diagnostics and inference
- Learn how to implement GLM in R

Reading:

- Chapter 3.5, 4.2, 4.4 of Dobson and Barnett

Homework: Assignment 1

Lecture 5 – Introduction of Logistic Regression

Learning Objectives:

- Learn binary response variables
- Understand the relation between grouped and ungrouped data
- Learn different link functions
- Build logistic regression and interpret model parameters

Reading:

- Chapter 7.1, 7.2, 7.4 of Dobson and Barnett

Homework: Assignment 2

Lecture 6 – Logistic Regression Inference and Diagnostics

Learning Objectives:

- Learn model inference and diagnostics
- Know how to use R to conduct logistic regression analysis

Reading:

- Chapter 7.3, 7.5, 7.6 of Dobson and Barnett

Homework: Assignment 2

Lecture 7 – Over-Dispersed Logistic Regression

Learning Objectives:

- Understand the source of over dispersion
- Learn to fit logistic models with over dispersion

Reading:

- Chapter 3.4 of Faraway

Homework: Assignment 3

Lecture 8 – Logistic Regression Case Study

Learning Objectives:

- Implement logistic regression in R
- Interpret model outputs

Reading:

- Chapter 7.8 of Dobson and Barnett

Homework: Assignment 3

Lecture 9 – Prospective/Retrospective Studies

Learning Objectives:

- Logistic model for prospective/retrospective studies

Reading:

- Chapter 4.3 of Faraway

Homework: Assignment 3

Lecture 10 – Multinomial Logistic Regression

Learning Objectives:

- Understand key properties of multinomial distribution
- Use multinomial logistic regression to model nominal categorical data
- Interpret model parameters
- Use statistical software for model fitting and inference in real data examples

Reading:

- Chapter 8.1, 8.2, 8.3 of Dobson and Barnett

Homework: Assignment 4

Lecture 11 – Ordinal Regression

Learning Objectives:

- Use proportional odds model and continuation ratio model to model ordinal multinomial data
- Interpret model parameters
- Use statistical software for model fitting and inference in real data examples

Reading:

- Chapter 8.4, 8.5 of Dobson and Barnett

Homework: Assignment 4

Lecture 12 – Poisson Regression

Learning Objectives:

- Review Poisson distribution properties
- Use statistical software to fit log linear models for count data

- Conduct model diagnostics and inference

Reading:

- Chapter 9.1, 9.2 of Dobson and Barnett

Homework: Assignment 5

Lecture 13 – Poisson Regression Extensions

Learning Objectives:

- Understand the phenomenon of over-dispersion in Poisson models
- Check over dispersion, and estimate dispersion parameter
- Learn zero-inflated Poisson regression model

Reading:

- Chapter 5.2, 5.5 of Faraway

Homework: Assignment 5

Lecture 14 – Contingency Table

Learning Objectives:

- Construct two-way contingency table
- Apply log linear models to two-way contingency tables

Reading:

- Chapter 9.3, 9.5 in book by Dobson and Barnett

Homework: Assignment 5

Lecture 15 – Recap of GLM

Learning Objectives:

- Review GLM
- Go through additional case studies

Reading:

Homework: Assignment 5

Lecture 16 – Mid-Term Exam

Lecture 17 – Introduction to Longitudinal Analysis

Learning Objectives:

- Recognize repeated measurements
- Understand basic concepts of longitudinal analysis

Reading:

- Chapter 1, 2 of Fitzmaurice, Laird and Ware

Homework: Assignment 6

Lecture 18 – General Linear Regression

Learning Objectives:

- Model mean and covariance of longitudinal data

- Learn restricted maximum likelihood

Reading:

- Chapter 3, 4.2, 4.5 of Fitzmaurice, Laird and Ware

Homework: Assignment 6

Lecture 19 – Linear Mixed Effects Model

Learning Objectives:

- Distinguish fixed and random effects
- Learn model estimation
- Use BLUP to predict random effects

Reading:

- Chapter 8.2, 8.3, 8.6 of Fitzmaurice, Laird and Ware

Homework: Assignment 7

Lecture 20 – Marginal Model: GEE

Learning Objectives:

- Review GLM
- Construct marginal model for longitudinal data with non-Gaussian response
- Learn generalized estimation equation (GEE) for model estimation

Reading:

- Chapter 12.1, 12.2, 12.3, 13.2 of Fitzmaurice, Laird and Ware

Homework: Assignment 8

Lecture 21– Generalized Linear Mixed Effects Model

Learning Objectives:

- Construct random effects model in GLM
- Learn model estimation and inference
- Contrast marginal and random effects models

Reading:

- Chapter 14.2, 14.3, 14.5 of Fitzmaurice, Laird and Ware

Homework: Assignment 8

Lecture 22 – Case Study

Learning Objectives:

- Apply LMM and GLMM to real data examples
- Interpret model parameters

Reading:

- Chapter 13.4, 14.7 of Fitzmaurice, Laird and Ware

Homework: Assignment 8

Lecture 23 – Introduction to Survival Analysis

Learning Objectives:

- Learn basic concepts in survival analysis (e.g., hazard, survival functions)
- Understand censoring mechanisms

Reading:

- Chapter 1 of Hosmer, Lemeshow and May

Homework: Assignment 9

Lecture 24 – Survival Function

Learning Objectives:

- Learn parametric models for survival functions
- Know how to calculate Kaplan-Meier curves

Reading:

- Chapter 2.2, 2.3 of Hosmer, Lemeshow and May

Homework: Assignment 9

Lecture 25 – Comparison of Survival Functions

Learning Objectives:

- Learn hypothesis testing between multiple survival functions

Reading:

- Chapter 2.4 of Hosmer, Lemeshow and May

Homework: Assignment 9

Lecture 26 – Cox Proportional Hazards Model

Learning Objectives:

- Build Cox model for survival data
- Learn parameter estimation
- Interpret model parameters

Reading:

- Chapter 3.1, 3.2, 3.3, 3.5 of Hosmer, Lemeshow and May

Homework: Assignment 10

Lecture 27 – Cox Model Extensions

Learning Objectives:

- Build Cox model for time-dependent covariates
- Cox model for stratified data

Reading:

- Chapter 7.2, 7.3 of Hosmer, Lemeshow and May

Homework: Assignment 10

Lecture 28 – Final Review

Learning Objectives:

- Review GLM, longitudinal analysis and survival analysis

Reading:

- None

Homework: None

ASSIGNMENTS

Assignment 1	
L2-L4	Assignment 1 covers basics about the exponential family distributions and GLM. Students will recognize the general form of the exponential family distributions, and select appropriate link functions for GLM.
Assignment 2	
L5-L6	Assignment 2 covers logistic regression. Students will use fit logistic models, interpret model parameters, and conduct inferences.
Assignment 3	
L7-L8	Assignment 3 contains a more comprehensive case study involving logistic regression.
Assignment 4	
L9-L10	Assignment 4 covers multinomial regressions. Students will select appropriate models for real data problems, fit multinomial logistic models and interpret model parameters.
Assignment 5	
L11-L13	Assignment 5 covers Poisson regression models. Students will fit Poisson models, with offset and over-dispersion, to count response data. The assignment will test model fitting, diagnostics, and parameter interpretation using real data.
Assignment 6	
L15-L16	Assignment 6 covers basics about longitudinal data analysis. Students will recognize repeated measures and know different types of covariance structures.
Assignment 7	
L17-L18	Assignment 7 covers mixed effects models. Students will build mixed effects models for real data problems, fit the models using statistical software and interpret model fitting results.
Assignment 8	
L19-L21	Assignment 8 covers GEE and GLMM. Students will implement GEE and GLMM to solve real data problems.
Assignment 9	
L22-24	Assignment 9 covers basics of survival analysis. Students will recognize different censoring mechanisms, understand hazard and survival functions, and be able to estimate and compare survival functions.
Assignment 10	
L25-27	Assignment 10 covers Cox proportional hazards model. Students will be able to fit Cox models and interpret model parameters for real data problems.