CLASS SESSIONS

Tuesday, 17:30 – 20:20 EST
Zoom: https://columbiacuimc.zoom.us/j/96800425701 Passcode: P8157-001

Time Slot (EST)	Description
17:30 - 18:00	Mini Exam slot on days when mini-exam is scheduled
18:00 - 20:20	Online Discussion

INSTRUCTOR

Srikesh Arunajadai, Ph.D.
sa2658@cumc.columbia.edu

TEACHING ASSISTANT(S)

Teaching Assistant	Email	Office Hours	URL (passcode)
Zilan Chai	zc2326@cumc.columbia.edu		
Muhire Kwizera	mhk2159@cumc.columbia.edu		
Anja Shahu	as6798@cumc.columbia.edu		

COURSE DESCRIPTION

Say that a vector of measurements of a certain variable of interest indexed by time is available for a set of units. Examples include

- Blood pressure measurements in every trimester in pregnant women
- Measurement of CD4 T-cell counts in men
- Measurements of pollen count from March to June on randomly selected day across 20 urban centers

The above examples from clinical, epidemiological, and environmental health sciences are candidates for Longitudinal Data Analysis. Longitudinal Data Analysis deals with the analysis of repeated measurements over time of a certain variable (blood pressure, CD4-T cell count, Pollen count) across a set of units (pregnant women, men, urban centers). As in linear regression models and generalized linear models. Often the question of interest is to understand the association between certain covariates of interest and outcome. A distinguishing feature in the analysis of Longitudinal data is the need to account for correlation between the outcomes from the same entity. The inference based on standard regression models is invalid for longitudinal data. In this course we will explore various methods to analyze longitudinal data with a special emphasis on applications to problems in public health and the biological sciences. The methods are easily portable to other areas of study with a similar data structure.

This is a required course for students in the PhD program in the department of Biostatistics and is a possible area of examination for both Inference and Applications part of the qualifying exam. In view of this, there will substantial and equal emphasis on both theory and applications throughout the course.

PREREQUISITES

- Calculus and Linear Algebra
- A reasonable understanding of the materials in P8104, P8109, and P8111
- Statistical analysis and computing will be an integral part of this course. Students are expected to have more than just the basic knowledge of a statistical software of their choice. This course will use R.

You may use a software of your choice but the neither the Tas or I will be able to help in softwares other than R. Note: This is not a course on statistical computing or programming. Lectures will deal with "how to approach the analysis" rather than "how to use the software for analysis". Code snippets will be provided if it deemed that particular tasks are of non-standard nature.

- If you are unsure of the above three – permission of the instructor

COURSE LEARNING OBJECTIVES

Students who successfully complete this course will be able to:

- Identify problems where the analysis of the data using Longitudinal data Analysis methods is viable and most appropriate
- Perform various exploratory data analysis and produce plots to understand and identify the key characteristics of a given longitudinal data
- Recognize the limitations of some classical methods to longitudinal data analysis
- Differentiate between marginal, random-effects, and transition models
- Analyze both Gaussian and non-Gaussian data
- Analyze longitudinal data with missing values
- Analyze longitudinal data with time dependent covariates
- Explore, identify, and specify models for the given data and perform model diagnostics using statistical software

ASSESSMENT AND GRADING POLICY

Student grades will be based on:

Homework (H) [There will 4 homework assignments].....[4 x 5] 20%
Mini-Exams (M) [There will 4 mini exams].....[4 x 10] 40%
Final (F).....[1 x 25] 25%
Final Oral (O).....[1 x 15] 15%

Scores will be computed:

S: based on above weights – 0.20 * H + 0.40 * M + 0.25 * F + 0.15 * O

Grading

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

- **Homework and Mini-exams**: Homework assignments are due as follows and will be assigned 2 weeks prior to the due date.

Date	Mini Exam	Homework Due
Sep 27	Mini Exam 1	HW 1
Oct 11	Mini Exam 2	HW 2
Nov 1	Mini Exam 3	HW 3
Nov 22	Mini Exam 4	HW 4

    - The assignments will comprise of a mix of analytical problems and problems related to use of a statistical software. The analytical problems are intended to enhance the understanding of the theory and assumptions related to the various concepts. The computing part is intended to enable students to put theory into practice.
    - Homework should be submitted through Courseworks
    - Assignments are due before end of the day (11:59 pm) on the due date
    - No late homework assignment submissions
    - While collaborative study habits are encouraged, blind copying is strictly prohibited. Students are encouraged to be familiar with the honor code of academic integrity provided below.
- **Mini exams**: These are 30 minutes online mini-exams to make sure you keep pace with weekly lessons. These exams are held on dates shown above at the start of the regular class. **No makeup mini-exams will be given**.
- **Final**: Online Final will be held on December 6, 2022. **No makeup final will be given**.
- **Final Oral:** A 10-minute oral exam conducted by the instructor or one of the TAs will be held on December 13, 2022. **No makeup exam will be given.**
- **Important Dates**

Task	Important Dates
Homework 1 and Mini-exam 1	September 27, 2022
Homework 2 and Mini-exam 2	October 11, 2022
Homework 3 and Mini-exam 3	November 1, 2022
Homework 4 and Mini-exam 4	November 22, 2022
Final Exam	December 6, 2022
Final Oral Exam	December 13, 2022

COURSE STRUCTURE

Recommended Textbooks:

- Applied longitudinal analysis, Garrett M Fitzmaurice, Nan M Laird, and James H Ware, Wiley; 2nd edition
- Analysis of longitudinal data, Peter J Diggle, Patrick Heagerty, Kung-Yee Liang, and Scott L Zeger; OUP Oxford; 2nd edition
- Linear mixed models for longitudinal data, Geert Verbeke and Geert Molenberghs; Springer; 1st ed.

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Students and faculty have a shared commitment to the School's mission, values and oath. Mailman.columbia.edu/about/mission-history

Academic Integrity

Students are required to adhere to the Mailman School Community Standards and Conduct handbook, which includes the Code of Academic Integrity.

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 proper notification of your recommended accommodations by Sarah Tooley(st3146@cumc.columbia.edu), Associate Director for Student Support and the School's liaison to the Office of Disability Services.

Course Content

2021			Contents
6-Sep	1	1.0	Overview
		1.1	Introduction
		1.2	Basic Notation
		1.3	Sources Of Correlation
		1.4	Why Model Correlation
		1.5	Merits Of Longitudinal
		1.6	Exploratory Data Analysis
13-Sep	2	2.0	Overview
		2.1	Matrix Essentials 1
		2.2	Matrix Essentials 2
		2.3	Multivariate General Linear Model
		2.4	Profile Analysis
		2.5	Multivariate Normal
		2.6	Wishart Distribution
		2.7	General Hypothesis Testing
20 - Sep	x		CATCH UP WEEK. NO NEW MATERIALS
27-Sep	3	3.0	Overview
		3.1	Linear Model
		3.2	Maximum Likelihood
		3.3	Inference
		3.4	REML
4-Oct	4	4.0	Overview
		4.1	GLM Review
		4.2	GEE
		4.3	Correlation 1
		4.4	Correlation 2

		4.5	Correlation 3
		4.6	Binary Example
11-Oct	5	5.0	Overview
		5.1	Quasi Likelihood
		5.2	GEE Estimation
		5.3	GEE Properties
		5.4	Sandwich Estimator
		5.5	Hypothesis Testing
		5.6	GEE Binary Revisited
		5.7	GEE Poisson
18-Oct	6	6.0	Overview
		6.1	Mixed Effects Intro
		6.2	Mixed Effects Setup
		6.3	Random Effects Prediction
		6.4	Linear Mixed Example
25-Oct	7	7.0	Overview
		7.1	LME To GLME
		7.2	GLME Setup
		7.3	GLME Interpretation
		7.4	Estimation
		7.5	Binary Example
		7.6	Poisson Example
1-Nov + 15-Nov	8 + 9	8.0	Overview
		8.1	Transition Model
		8.2	Transition Example
		8.3	Time Varying Covariates
		8.4	Full Partly Conditional Means
		8.5	Lagged Covariates
		8.6	Endogeneity
		8.7	Counterfactuals
		8.8	Endogeneity Estimation
22-Nov + 29-Nov	10 + 11	9.0	Overview
		9.1	Concepts In Incomplete Data
		9.2	Longitudinal Implications
		9.3	Dropout
		9.4	Handling Missing Data