

COURSE SYLLABUS
BIostatISTICS 755–SPRING 2023
INTRODUCTION TO LONGITUDINAL AND MULTILEVEL MODELING
10:05:11:20 TUESDAY/THURSDAY
PUBLIC HEALTH RESEARCH CENTER 320

- **Instructor:** Alexander McLain, PhD. E-mail: mclaina@mailbox.sc.edu, Office: Discovery I Room 450, Office Phone: (803)777-1124, Fax: (803)777-2524.

– **Office Hours:** Monday 10:30–12:00 and by appointment.

- **Teaching Assistants:**

– Peter Olejua, polejua@email.sc.edu
– Md. Nasim Saba Nishat, mnishat@mailbox.sc.edu.

- **Course Website:** The course website can be found [here](#). This is where all notes and homework's will be posted, please bookmark it.

- **Class Communication:** We will use [Slack](#) as a discussion board throughout the semester. Please use this to ask questions about homework or other course topics. This will be regularly monitored by the course instructor and teaching assistants. All questions will be addressed within 24 hours of posting.

If there are homework questions you are not comfortable posting on Slack you may e-mail the instructor or TAs. These questions will be redirected to Slack and answered in due course. For questions about your projects e-mail is the preferred method of communication, however, if the question is general enough it will be reposted on Slack. Invitations to our Slack channel will be sent to your school email.

- **Main Reference:** Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (2nd Ed.) John Wiley & Sons. Hereafter FLW. The website for this text, which contains datasets and sample SAS code, can be found [here](#).

- **Other Texts (not required):**

– Gelman, Andrew, and Jennifer Hill (2014). *Data analysis using regression and multilevel/hierarchical models. Vol. 1.* New York, NY, USA: Cambridge University Press. Hereafter GH.
– Pinheiro, J., and Bates, D. (2006). *Mixed-effects models in S and S-PLUS.* Springer Science & Business Media.
– Hedeker, D. and Gibbons, R.D., (2006). *Longitudinal data analysis.* John Wiley & Sons.

- **Course Description:**

Longitudinal data consist of multiple measures over time on a sample of individuals. This type of data occurs extensively in both observational and experimental biomedical and public health studies, as well as in studies in sociology and applied economics. This course will provide an introduction to the principles and methods for the analysis of

longitudinal and multilevel data. Longitudinal data are a special case of clustered data, where observations are clustered within people. Multilevel modeling generalizes the ideas we'll learn about longitudinal data to the case where the clustering can have multiple levels (e.g., students within classrooms within schools etc.). Primarily interest will be in finding the answer to a specific research question, and not statistical theory.

We will focus on how to identify an appropriate statistical technique, apply the technique in SAS, diagnose the fit of the model, and interpret the results. Emphasis will be on data analysis, interpretation, and case studies. Problems will be motivated by applications in epidemiology and clinical medicine, health services research, and disease natural history studies.

Pre-requisites: B or better in BIOS 757 or similar course. BIOS 757 is an introduction to linear and logistic regression.

- **Learning Outcomes:** A student who successfully completes this course will:
 - know which longitudinal/multilevel statistical models and methods to apply to a particular data example,
 - identify an appropriate missing data strategy,
 - develop a covariate strategy that results in the hypotheses tests of interest,
 - implement the appropriate technique in statistical software,
 - assess the fit of the model based output,
 - have the ability to explain the meaning of the estimated coefficients, hypotheses tests and confidence intervals to a broad audience.
- **Course Work:**
 - *Homework (60%)* We will typically have 5-6 homework assignments. You are allowed to work together on homework assignments, as long as your homework does not become a copy of another. These will generally be due the next week, and will be graded and returned at the class that follows. Homework assignments turned in after 5:00 PM on the date due, but before the return date, will have the grade reduced by 20%. Homework assignments will not be accepted after the return date, unless arrangements have been made with the instructor prior to the due date, or because of exceptional circumstances.
 - *Mid-term project (15%)*
 - *Final project (25%)* See the *Project Description* handout for more information on the projects.

You will work in pairs for both of the projects.

Grades will be assigned as follows: 91–100= A ; 88–91= B^+ ; 81–88= B ; 78–81= C^+ ; 71–78= C , 62–71= D , and 0–62= F .

- **Cheating:** No form of cheating or plagiarism will be tolerated in any way. Anything that you cheat on (test, homework, or project), you will receive no credit for. A copy of the Carolinian Creed is available to all students; this should be read and understood.

- **SAS:** Many students find it more convenient to purchase SAS for their home computers. SAS is currently not available on the iOS platform. If you must use iOS most will opt for the SAS OnDemand for Academics which is available via welcome.oda.sas.com. If you choose to purchase SAS you can do so through the university at a much better price. SAS is also available at many of the computer labs on campus including the 4th floor of Discovery I.

Some SAS tips:

- Get comfortable with SAS. There are countless online videos about how to use the data step.
- We will not do anything in the way of “coding” in SAS. All datasets will be prepared for you, but this is not how it is in the real world.
- SAS has extensive online documentation and examples.
- SAS doesn’t break.
- SAS does exactly what you tell it to do, even if it is wrong.

I will try to post companion examples that use R when possible. These are from when I used R as the main software in Spring 2020. I stopped using R as the main software due to it’s limitations for fitting for multinomial and negative binomial distributions, and the computational limitations GLMMs. I do not recommend using STATA.

- **Academic Integrity:** You are expected to practice the highest possible standards of academic integrity. Students may brainstorm ideas for homework assignments, but may not copy solutions from other students or from other sources. Any deviation from this expectation will result in a minimum academic penalty of your failing the assignment, and may result in additional disciplinary measures. This includes improper citation of sources, using another student’s work, and any other form of academic misrepresentation.
- **Attendance Policy:** Though attendance is not required, it is strongly recommended.
- **Disability Resource Center:** The [Student Disability Resource Center \(SDRC\)](#) empowers students to manage challenges and limitations imposed by disabilities. Students with disabilities are encouraged to contact me to discuss the logistics of any accommodations needed to fulfill course requirements (within the first week of the semester). To receive reasonable accommodations from me, you must be registered with the Student Disability Resource Center (1705 College Street, Close-Hipp Suite 102, Columbia, SC 29208, 803-777-6142; email: sadrc@mailbox.sc.edu). Any student with a documented disability should contact the SDRC to arrange for appropriate accommodations.
- **Course Outline:**

Week	Goals	Chapters*
1	Introduction, Basic Concepts, Data Examples, Simple approaches	1-2
2	Random effects ANOVA, time as a covariate, MANOVA	3
3	Growth curve modeling, interactions over time	4–6
4	Covariance Pattern Analysis Linear Mixed Models	7
5	Linear Mixed Models	8
6	Fixed versus Random effects and Model diagnostics	9–10
7	Generalized Linear Models, Generalized Estimating Equations (GEE)	11–12
8	Mid-term presentations	
	Spring Break	
9	GEE and Generalized Linear Mixed Models (GLMM)	13-14
10	GLMM vs GEE, Missing Data in Longitudinal Studies	9, 16
11	Missing Data (cont.), Multi-level Introduction and Data Structure	17
12	Multi-level Linear Models	11 GH
13	Multi-level Growth Curves and GLMs	12–15 GH
14	Final Presentations	

*text is FLW unless noted otherwise.