Course Syllabus
Biostatistics 755–Spring 2024
Introduction to Longitudinal and Multilevel Modeling
10:05:11:20 Tuesday/Thursday
Swearingen 2A22

- **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:**
  - Robert Aidoo, Biostatistics PhD Student. Email:RAIDOO@email.sc.edu. Office hours: Tuesdays 4-5pm in Discovery 435A.
- **Course Website:** The course website can be found here. This is where all notes and homework 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 assistant. 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 TA. 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 e-mail. If you do not receive one by the end of the week please e-mail the TA's.
- **Main Reference:** Hedeker, D. and Gibbons, R.D., (2006). *Longitudinal data analysis*. John Wiley & Sons. 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.
  - 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.
  - Pinheiro, J., and Bates, D. (2006). *Mixed-effects models in S and S-PLUS*. Springer Science & Business Media.
- **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 introduce the principles and methods for analyzing 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 finding the answer to a specific research question, not statistical theory.

We will focus on identifying an appropriate statistical technique, applying the technique in SAS, diagnosing the model's fit, and interpreting the results. Emphasis will be on data analysis, interpretation, and case studies. Problems will be motivated by epidemiology and clinical medicine applications, 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,
  - implements the appropriate technique in statistical software,
  - assess the fit of the model-based output,
  - has the ability to explain the meaning of the estimated coefficients, hypotheses tests, and confidence intervals to a broad audience.
- **Course Work:**
  - *Homework (60%)* We typically have 5–6 homework assignments. You can work together on homework assignments if your homework does not become a copy of another. Assignments will be due by 5:00 pm on the assigned day, which will be a weekday (no midnight or Sunday due dates). They will be graded and returned by the following week. Homework assignments turned in after 5:00 PM on the date due but before the return date will have the grade reduced by 10%. 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.

    The homework should be e-mailed to both the TA and the course instructor. **When emailing your solutions use "last-first name-HW#" for all files.** For example, if I am handing in a Word and SAS file for the first homework, they would be titled "McLain-Alex-HW1.docx" and "McLain-Alex-HW1.sas". Do not feel that you have to format the homework (i.e., cutting and pasting the output is sufficient), but the answers to your questions should be clear.
  - *Mid-term project (15%)*
  - *Final project (25%)*

See the *Project Description* handout for more information on these projects. Grades will be assigned as follows: 90–100=A; 87–90=B^+; 80–87=B; 77–80=C^+; 70–77=C, 62–70=D, and 0–62=F.

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

- **SAS:** Many students find purchasing SAS for their home computers more convenient. 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 static and cleaned.
  - SAS has extensive online documentation and examples.
  - SAS doesn't break. It 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 its limitations in fitting multinomial and negative binomial distributions and computational limitations with 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 for 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	Covariance Pattern Analysis, time as a covariate	6
3	Growth curve modeling, interactions over time	2
4	Random effects ANOVA, MANOVA	2 - 3
5	Linear Mixed Models	4
6	Fixed versus Random effects and Model diagnostics	9–10 FLW
7	Generalized Linear Models, Generalized Estimating Equations (GEE)	8
8	**Mid-term presentations**	
	Spring Break	
9	GEE and Generalized Linear Mixed Models (GLMM) for Binary	8–9
10	GLMM for multinomial and count outcomes	10–12
11	Missing Data, Multi-level Introduction and Data Structure	14
12	Multi-level Linear Models	11 GH
13	Multi-level Growth Curves and GLMs	12–15 GH
14	**Final Presentations**	

*text is HG unless noted otherwise.