BIOS 663: Intermediate Linear Models

Instructor:

Fei Zou

4115C McGavran Greenberg Hall

Email: feizou@email.unc.edu

Tel: 919-843-4822

Class: 1:25-3:10 PM MW (MC 1305)

Lab Session: 3:35-4:35 PM M (MC 1305)

Office Hours: 1 AM - 2:30 PM T or by appointment

Texts:

- Muller and Fetterman, *Regression and ANOVA: An Integrated Approach Using SAS Software* (required)
- Weisberg, *Applied Linear Regression, 4th Edition* (required). Available online through UNC libraries. R supplement available at

  `http://users.stat.umn.edu/~sandy/alr4ed/links/alrprimer.pdf`
- Kleinbaum, Kupper, Muller, and Nizam, *Applied Regression Analysis and Other Multivariable Methods* (recommended)
- Harrell, *Regression Modeling Strategies* (recommended)

BIOS 663 offers a matrix-based treatment of general linear regression, generalized linear regression, and experimental design with applications in the health sciences. Emphasis is placed on hypothesis testing, diagnostics, and model building, as well as on interpretation and communication of statistical results. The course begins with a review of matrix algebra and concludes with an introduction to generalized linear models.

In BIOS 663, students will

1. Learn to analyze and interpret univariate linear models,
2. Gain a basic understanding of the associated theory of linear models in order to know when not to apply the methods and how to extend the theory to non-standard situations,
3. Gain exposure to more advanced models, including generalized linear models,
4. Gain basic knowledge in experimental design, and
5. Learn to communicate statistical results to subject-area collaborators (including public health professionals, physicians, and other scientists and researchers) in non-statistical language so that the results can be understood.

More specifically, we will touch on the following topics, subject to possible changes.

1. Linear Algebra Review.
2. Introduction to R
3. The Linear Model: Estimation and Testing
4. Some Distributional Results for the LM
5. Multiple Regression: General Consideration
6. Testing Hypotheses in Multiple Regression
7. Correlation
8. Assumption Diagnostics
9. Computation Diagnostics
10. Polynomials and Splines
11. Transformations
12. Selecting the Best Model
13. Coding Schemes for Regression
14. Logistic Regression
15. Poisson Regression

16. Experimental design, power and sample size calculation

Course Prerequisites: The following prerequisite skills are assumed.

- Working knowledge of calculus and matrix algebra (through MATH 547)
- Working knowledge of probability and inference (through BIOS 550/660)
- General applied statistics knowledge (at the level of BIOS 662)
- Working knowledge of SAS (at the level of BIOS 511) or R

Students are responsible for ensuring they have the necessary prerequisites. Violations of these assumptions will likely lead to frustration on the part of the student! BIOS 663 is primarily for students in the biostatistics department, though quantitative-minded students from other departments are welcome (if you are such a student, see me if you have questions about your mathematical background). The department also offers two other courses in linear models: BIOS 762 is an advanced theoretical treatment of linear models, and BIOS 545 is a basic applied course in linear models. Students without the necessary prerequisites (MATH 547, BIOS 511, BIOS 550 or BIOS 660, BIOS 662) should consider other courses in the BIOS department or in the school.

Copies of slides used in class will be available online at

`https://sakai.unc.edu/portal/site/bios663_2019`

These slides cover some material not contained in the texts and do not cover all material in the texts. You will be responsible for topics only covered in lecture. You are encouraged to print the slides and bring them with you to class.

Graded Work : Grades will be determined as follows:

- Homework assignments and data analysis project: 25%
- Midterm exam: 35%
- Final exam: 40%

Graduate courses in the School of Public Health use the following grading system.

- **H**: Clear excellence
- **P**: Entirely satisfactory
- **L**: Low passing
- **F**: Fail

All assignments are cumulative. The SPH grading system is designed so that the mode of the grading distribution is **P**.

Descriptions of Graded Work:

- **Homework and data analysis project** Homework problems are designed to ensure that material from the text and course notes has been understood and mastered. Homework will be given approximately biweekly, and students are allowed to talk about ideas and approaches to problems in groups, though students should "write up" assignments independently. It is to the student's benefit to attempt these problems on their own initially as practice for later in-class exams. The data analysis project will be a group project and you are expected to apply concepts learned in this class to one of real data sets.

- **Exams** There will be one in-class midterm exam (March 6th, tentatively) and one final exam (Thursday May 2nd at 12PM). The exams are designed to test your mastery of the material presented, your ability to conduct statistical analyses on your own, and your ability to interpret the results of analyses in language that subject-area investigators can easily understand.

The Student Honor Code: Students are encouraged to work together on homework, but copying someone else's work *always* creates an honor code violation. Students are never allowed to discuss or work together on exams while they are being given. Expulsion from the university is possible if the honor code is violated, and receiving 0% on the assignment in question is a certainty.