STA 6246 Linear Models
Fall 2020

Instructor Hani Doss—222 Griffin-Floyd; email: doss@stat.ufl.edu; Office Hours: They will be MWF period 3 (9:35am–10:25am), and you may either call me on my cell phone or talk to me via zoom (you may also email me, although asking technical questions by email often doesn't work, and in any case you should not expect an immediate response). However, depending on the situation it may be more convenient to set up a personal zoom meeting at a different time. I will email the class the following information, which you should not give out to anyone who is not in the class: my cell phone number; my zoom personal ID; and the username and password for the parts of the course webpage that are password protected.

Teaching Assistant Maoran Xu—email: maoranxu@ufl.edu.

Course Description This course covers the theory of the general linear model, including regression and analysis of variance models, least squares estimates and distribution theory when the errors are normally distributed. A geometric approach is emphasized.

Prerequisites Prerequisites are:

- STA 6208 Basic Design and Analysis of Experiments
- STA 6327 Introduction to Theoretical Statistics II

If you have not taken both of these (or their equivalents), you may not register for this course.

Orientation of the Course The primary purpose of this course is to prepare the students in the Ph.D. program in Statistics for the linear models portion of the Ph.D. qualifying exam in Statistics. As such, I will cover a specific body of material in the theory of linear models that forms the core of what every Ph.D. student in Statistics should know. This is not a course on experimental design, nor is it a course on applied regression. If what you are looking for is a course on these topics, you should drop this course immediately.

Text R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models* (5th edition, 2020), Springer. (The fourth edition is OK also.)

We will use the statistical computing language R (which can be downloaded for free from http://www.r-project.org), and a student who is not familiar with it is strongly advised to become so by the fourth week of the semester.

We will not use the book by Christensen much. However, it is an excellent text, and the approach used in that book is close to the approach we will use. You may not wish to buy it, but you need to be advised that every serious doctoral student in Statistics must own a good book on linear models.

Grading Your final course grade will be based on the four components below, with the stated weights:

Exam 1:	Wednesday October 7, 8:20 pm; covers everything up to and including the lecture of Monday October 5. Note the evening time slot.	25%
Exam 2:	Friday November 13, 8:20 pm; covers all material after Exam 1 up to and including the lecture of Mon November 9. Note the evening time slot.	25%
Final:	Comprehensive. Friday December 18, 7:30 am–9:30 am.	35%
HW:	About 8 or 9 homework assignments.	15%

The solutions to the homework assignments must be entirely your own (this applies also to R code).

For Exams 1 and 2, you will have two hours for a 50-minute exam.

Initial Assignment

1 Read Appendix A of PACQ. You need to have read the first half before the second lecture.

2 Do all the exercises in Appendix A (but for exercises with long lists only do a few parts). This is not to be handed in. Please do this before the third lecture.

3 Read pages 13–22 of the class notes. Please do this before the third lecture. Note: some of this material duplicates, or elaborates on, what is in Appendix A of PACQ.

4 Start to get familiar with R if you aren't already.

Note: Items 1–3 above should be a review.

Course Web Page `http://www.stat.ufl.edu/~doss/Courses/lm`

Topics

- Overview of general linear models.
- Review of basic linear algebra, including projections, eigenvalues and eigenvectors, and spectral decomposition of symmetric matrices.
- The multivariate normal distribution.
- Least squares estimates: their derivation and basic properties for models of full rank. Relationship between the normal equations and projection operator approaches. Weighted least squares.
- Distribution of least squares estimates under normal errors.
- One-way and two-way analysis of variance.
- Linear regression.
- Implementation in R.

- Introduction to nonparametric regression (local linear smoothing, regression splines, smoothing splines); generalized additive models; implementation of these in R.
- Hypothesis testing. The general linear hypothesis. Discussion of full-rank and less-than-full-rank models. The F test; the likelihood ratio test and connections with the classical tests in large sample theory for parametric models.
- Penalized regression, including ridge regression and lasso.
- Multiple comparisons.
- Bayesian analysis of linear models.

If you need special arrangements because of a disability please see me.

PLAN FOR ONLINE INSTRUCTION

Lectures I will use Zoom. Lectures will be asynchronous, meaning non-live, so you will not be able to ask questions during the lecture (you can ask questions during office hours of course), but you will be able to watch the lecture any time you want. Every time I create a zoom lecture, I will install it on our course web page in the folder Videolectures-and-notes and send you an email to that effect. The basis for each video lecture will be a set of slides, which will be considerably more detailed than the slides I make for the face-to-face version of the course, to take into account that I will not have two blackboards. In each video lecture, I will go over the slides, adding explanations and details. I will install in Videolectures-and-notes a pdf file containing the slides. You should print the pdf file prior to watching the video, so that you can write notes on it.

Homework Every time I create a homework assignment I will notify you by email. You will download the assignment from the Homeworks folder on the course web page. Homeworks are to be emailed by noon on the day they are due. Late homework will not be accepted. The homeworks will typically have two parts, a handwritten part for the usual theoretical problems, and a report part for problems that involve R. You may write the solutions to the theoretical problems by hand on paper and then scan the paper. The solutions to problems that involve R should be typed in a pdf file. You should email the homeworks to the TA (Maoran Xu, `maoranxu@ufl.edu`), and not to me. She will email you the graded homeworks (with cc to me). Some time in the afternoon of the day on which the homework is due, I will install the solutions, and I will notify you by email when I have done so.

Exams The exams will be take home. I will email you the exam as a pdf file, you will print it, write your solutions on the printed file, sign a statement that you have not given nor received any help on the exam, scan the exam and email it to me (me, not the TA) within the two-hour limit. Please write very legibly, because scanning reduces resolution. I will email you the graded exam.