STA 6246 Linear Models
Fall 2016

Instructor Hani Doss, 222 Griffin-Floyd Hall; Office Hours: MWF 10:40-11:30 am (if you come to see me, please do so before 11:10), or by appointment; Phone: 273-2991; email: `dos s Øst at .ufi. edu` (do not ask questions regarding the course material by email).

Teaching Assistant Liyuan Zhang—115A Griffin-Floyd Hall; Office Hour: Thursday, 4th period (10:40-11:30 am); email: `zhangliyuanøuf 1. edu`

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

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

• STA 6329 Matrix Algebra and Statistical Computing

If you have not taken all of these, 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.

Texts (i) R. Christensen, *Plane Answers to Complex Questions: The Theory of Linear Models* (4th edition, 2011), Springer. (The third edition is OK also.)

(ii) W.N. Venables and B.D. Ripley, *Modern Applied Statistics with S* (4th edition, 2002), Springer. (Actually, any good book on R would be OK.)

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. This could be done for example by reading in Venables and Ripley Chapters 1-3 (skipping Section 3.10), Chapter 5, Sections 1-3, and taking a look at Chapter 4. If you prefer to use other statistical languages or statistical packages and do not intend to learn R, you should drop this course immediately.

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 September 28, 8:20 pm	28%
Exam 2:	Friday November 4, 8:20 pm	28%
Final:	Thursday December 15, 3:00-5:00 pm 36%	
Homework:	about 8 homework assignments	8%

Note the evening time slots for the first two exams. 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 1:40 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.

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

Topics

• Overview of general linear models.

• Review of basic linear algebra: projections; orthogonal decompositions; eigenvalue-eigenvector decomposition (spectral decomposition).

• The multivariate normal distribution.

• Feast 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.

• Finear 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.

- Multifactor analysis of variance.

- Bayesian analysis of linear models.

- Variable selection; collinearity; ridge regression; lasso and least angle regression.