Math 439 **Fall 2014**

LINEAR STATISTICAL MODELS

Welcome to Math 439! My name is Professor Edward Spitznagel. This is a course that provides the matrix theory (AKA linear algebra) that underlies regression and analysis of variance. You will find that neither of our standard courses in matrix theory, Math 309 and Math 429 contains quite the same material as what you will learn in this course.

That's because linear algebra is a veritable Swiss Army Knife, adaptable to and used in virtually all areas of mathematics. When I was a graduate student at University of Chicago, a student complained to his fellows about how tough the matrix theory course was. It so happened that the department chairman, Irving Kaplansky, was walking by, and the student didn't notice him. Kaplansky snuck up behind him and said in a booming voice, "**Everything is a matrix**." The student never forgot that and in fact enjoyed retelling it every chance he had.

So what's different about matrices in Math 439? For one thing, they all have real coefficients and have real eigenvalues. Lots of matrices are square ($n \times n$). Matrices are often not invertible, but they all have *generalized inverses*. Generalized inverses are closely connected with least-squares estimation and hypothesis tests. Certain square matrices generate the sums of squares in hypothesis tests. The ranks of those matrices are the "degrees of freedom" in the tests.

Textbook

The text is Ravindra Bapat's *Linear Algebra and Linear Models*. You can download it ***for free*** in pdf format through Olin Library at:

`http://catalog.wustl.edu:80/record=b4777305~S2`

When you download the text, you will also be given the opportunity to buy a printed copy for $24.99 and have it mailed to you. I have asked the bookstore to stock a few copies of the book, in case you find buying it from them to be more convenient. The bookstore price is $49.95 plus sales tax.

R Software

In keeping with the many uses of linear algebra, different software packages have been developed with different matrix capabilities. Some are commercial, and some are free. To name a few, there are: Matlab, Maple, Mathematica, Macsyma, Axiom, Derive, GAP, Sage, Scilab, Octave, SAS IML, Stata Mata, S, and R. These last two packages S and R have a common heritage, and it is unabashedly statistical. Some 40 years ago, Bell Labs developed three software innovations that shook the computing world: The operating system Unix, the compiler language C, and the statistics package S. The first two, Unix and C, have kept their names, while S evolved into R. Actually, it is still possible to buy S for a "mere" $1999, but R is free, and almost everyone has deserted S in favor of R.

Both R and S operate under the same philosophy, in which smart statisticians have contributed functions, and these functions do the heavy lifting. Most R users perform their work by using a few of these very powerful functions. For instance, to calculate a regression of y on x, they can evaluate the function lm(y~x), where "lm" stands for linear model. In Math 439, we will drop down one level and do the same thing with basic matrix operations, such as multiplication, transposition, and inversion: $(X'X)^{-1}X'Y$. (Well, this formula is a slight oversimplification, but the idea is sound, and it easily translates to R.)

Why use matrix functions?

First, if we limited ourselves to functions like lm(), R is simply a "black box," inside which we don't really know what's going on. That is, we wouldn't learn the theory behind linear statistical models.

Second, there are lots of advanced statistical methods that are defined in terms of matrix equations, and the only way to implement them is through matrix calculations. At one of the annual SAS Users Group meetings, John Sall, the vice president of SAS Institute, explained that SAS had become so popular they were unable to keep up with user requests for new statistical methods. They realized that the one thing these methods all had in common was that they were defined in terms of matrices. Therefore SAS Institute developed the Interactive Matrix Language (IML) procedure to allow their users to do their own matrix calculations. We will be doing the same thing with R, but without the "overhead" of learning SAS.

R Operators and Functions

In our course we will build virtually everything we need from several dozen R operators and functions. Documentation of many of these functions can be found at:

http://www.statmethods.net/management/operators.html

http://www.statmethods.net/management/functions.html

http://www.statmethods.net/advstats/matrix.html

You might like to print those several pages of documentation and keep them handy when you do your homework. In the first few homework assignments, I will take you by the hand and lead you through very explicit examples of how to use them (like the "case studies" of MBA programs). The explicit nature of these examples will gradually taper off as the course progresses, and you become more familiar with R and can work on your own and with each other.

Course Schedule

Our course meets Mondays and, Wednesdays, 4-5:30 in Cupples I Room 115. **Before you come to class, please preview the section or sections of the book to be covered that day.** Our schedule is as follows:

Day	Sections
Monday, August 25	1.1 TO 1.3
Wednesday, August 27	2.1 TO 2.3
Wednesday, September 03	2.4
Monday, September 08	3.1 TO 3.2
Wednesday, September 10	3.3
Monday, September 15	4.1 TO 4.2
Wednesday, September 17	4.3
Monday, September 22	5.1 TO 5.2
Wednesday, September 24	5.3 TO 5.4
Monday, September 29	5.5
Wednesday, October 01	6.1
Monday, October 06	6.2
Wednesday, October 08	7.1
Monday, October 13	7.2
Wednesday, October 15	MIDTERM EXAM
Monday, October 20	7.3 TO 7.4
Wednesday, October 22	8.1
Monday, October 27	8.2
Wednesday, October 29	8.3
Monday, November 03	8.4
Wednesday, November 05	8.5
Monday, November 10	9.1
Wednesday, November 12	9.2
Monday, November 17	9.3
Wednesday, November 19	9.4
Monday, November 24	10.1 TO 10.2
Monday, December 01	10.3
Wednesday, December 03	10.4 TO 10.5

Homework Schedule

With the exception of the first day of class (August 25th) and the class after the midterm exam (October 20th), a small amount of homework is due at the beginning of each class. The schedule is as follows:

Day	Exercises
Monday, August 25	Nothing due
Wednesday, August 27	1.1,1.3,1.5
Wednesday, September 03	2.1,2.6,2.8
Monday, September 08	2.10,2.15,2.19
Wednesday, September 10	3.2,3.6,3.12
Monday, September 15	3.15,3.18,3.27
Wednesday, September 17	4.1,4.3,4.6
Monday, September 22	5.8,5.10,5.11
Wednesday, September 24	5.1,5.3,5.4
Monday, September 29	5.8,5.10,5.11
Wednesday, October 01	5.12,5.14
Monday, October 06	6.1,6.4,6.5
Wednesday, October 08	6.6,6.8,6.11
Monday, October 13	7.1i,ii,iii
Wednesday, October 15	7.2,7.3,7.4
Monday, October 20	Nothing due
Wednesday, October 22	7.10,7.11,7.12
Monday, October 27	8.1,8.2,8.3
Wednesday, October 29	8.4,8.5
Monday, November 03	8.6,8.7,8.21
Wednesday, November 05	8.8,8.9,8.22
Monday, November 10	8.10,8.11
Wednesday, November 12	9.1,9.2
Monday, November 17	9.3,9.4

Wednesday, November 19	9.5,9.6
Monday, November 24	9.9,9.10
Monday, December 01	10.1,10.2,10.3
Wednesday, December 03	10.7,10.9

For the most part, two or three homework problems are required per class. They are due at the beginning of class. Three of you will then be randomly selected to present solutions to those problems on the blackboard, one problem per student. (To do so, you temporarily "borrow back" the work you just handed in.)

While this kind of classroom participation may be unusual in mathematics courses, it is very common in other departments. It encourages you to stay current in your studies.

My *official* office hours are from 5:30 to 6:30 on Monday and Wednesday. That's right after class. My office is Room 118 in Cupples I, just a few feet east of our classroom. As far as unofficial hours are concerned, you are *welcome* to knock anytime you see the light on. However, if you are elsewhere on campus or off-campus, I recommend calling in advance to see if I'm in. My telephone number is 935-6745.

Calculators

The **Texas Instruments calculators TI-83, TI-84, and TI-89 (and the new TI-Nspire series)** can perform some but not all matrix calculations. You may find them helpful in checking your work in some of the exercises. They will also be useful on your midterm and final exams. However, there will be nothing on your exams that requires their special capabilities. A scientific "four-banger" like a TI-30 should suffice. That is, you will need a calculator that can add, subtract, multiply, divide and (occasionally) take a square root, logarithm, or exponential. If you have a more powerful calculator made by TI, HP, Sharp, and Casio, etc., you are more than welcome to use it on the exams. In fact, feel free to bring a bunch of calculators if you wish. Just make sure you have new (or spare) batteries.

Accessing R

The R package is free and can be installed on PC's, Macs, and various flavors of Unix. To download and install it, go to http://cran.r-project.org/ and follow the instructions.

The R package is also "portable" in the sense that it can be installed on a USB flash drive. In fact, that is the mode I use for classroom teaching. If you like, I will let you copy my portable version of R onto a flash drive.

Finally, R is available on the 14 computers in Seigle Hall Room L012.

This course is not intended to turn you into skilled R programmers, and we pretty much limit ourselves to the matrix functions of R, and not its statistical procedures or data structures. However, if you would like to learn more about R, you can download Mike Allerhand's *A Tiny Handbook on R* through Olin Library at:

http://catalog.wustl.edu:80/record=b4637840~S2

There are dozens of other books on R available through Olin Library, mostly published by Springer-Verlag.

Examinations

We will have a midterm exam on October 15th and a final exam on December 12th. The midterm is at the regular class time, 4:00-5:30, while the final exam is 6:00-8:00. The midterm is in our regular classroom, Cupples I 115. The final exam may be in a different room, to be announced at the end of the semester. Each examination will consist of fifteen long-answer problems, each worth one point. You may bring one 8.5×11 inch notesheet to each examination. You may write on both sides of the notesheet. In the past, some students have formed "cooperatives" and composed their notesheets using MS Word. If you wish to do that, it's fine with me.

Course Grades

Most homework sets contain three problems each, while some contain just two. There are a total of 70 problems. Points are awarded in discrete fashion, 1 point if a solution is mostly correct and 0 if a solution is mostly incorrect. The two exams are worth a total of 30 points, so homeworks plus exams are worth 100 points. Correct presentations are worth up two extra points each. Typically I will invite you to present three times during the semester, so ideally you can accumulate up to 106 points by the end of the course.

I anticipate that the course grades will follow the modern convention, in which the A range will be 90 to 100, the B range will be 80 to 90, the C range will be 70 to 80, and the D range will be 60 to 70, with plus and minus grades at the tops and bottoms of each of these ranges. If you are registered pass/fail, you must achieve at least 70 points to pass, which is the lowest score for a C−.)