36-463/663: Topics in Statistics: Hierarchical & Multilevel Models
Fall 2016

TTh 10:30--11:50, DH A302

http://www.stat.cmu.edu/~brian/463-663

Course Information

Instructor:	*TA:*
Brian Junker, Statistics	Nick Kim
132E Baker Hall	FMS 328
(412) 268-8874	(412) 251-4637 (cell)
brian@stat.cmu.edu	nicolask@stat.cmu.edu
Office Hours (BH 132E):	*Office Hours:*
3-4pm Tue & Thu	TBA, See http://www.stat.cmu.edu/~brian/463-663/
Sometimes, Tue after class as well	*(or by appointment)*.
(or by appointment).	

Prerequisites

The formal prerequisite for the course is Probability Theory and Mathematical Statistics (36-225/36-226). In addition students should have taken, or be concurrently registered for, a high level linear regression course such as 36-401, 36-607 or 36-617.

One one hand, the course will be taught at a fairly high level—this means I actually expect to be able to talk to you about statistics at the level of 36-226 (calculus-based statistics) and 21-240 or 21-241 (matrix algebra). On the other hand, this course will have a very hands-on flavor. We are actually going to be doing computing and data analysis in R and other packages.

Please feel free to contact me if you have any questions or need additional information.

Texts

The required text for this course is

- Gelman, A. & Hill, J. (2007). *Data Analysis Using Regression and Multilevel/Hierarchical Models.* NY: Cambridge Univ Press. $63.74 in paperback at Amazon.com[1]

I also recommend that you find a copy of

- Lynch, Scott M. (2007). *Introduction to Applied Bayesian Statistics and Estimation for Social Scientists.* New York: Springer. $161.79 in hardcover at Amazon.com[2]

for later in the course. Most other materials for the course will be available on-line.

[1] Somewhere online there is pdf of the whole book too, but even if you find the pdf, you should buy the paperback book to support authors who write good textbooks at affordable prices!

[2] There used to be a pre-publication pdf of the book online at http://www.princeton.edu/~slynch/bayesbook.pdf, but it seems to have disappeared. I will try to make relevant material available in handouts, if you do not want to buy this book.

Course Description and Course Objectives

Multilevel and hierarchical models are among the most broadly applied "sophisticated" statistical models, especially in the social and biological sciences. They apply to situations in which the data "cluster" naturally into groups of units that are more related to each other than they are the rest of the data, and where relationships between variables are different for each individual or group of individuals.

My goals for you in this course are:

- To understand multilevel models as a generalization of linear regression models.
- To understand multilevel models as a kind of hierarchical Bayes model.
- To gain experience building, fitting and evaluating multilevel and hierarchical models, using classical and Bayesian methods.
- To gain experience analyzing data in a variety of settings in which these models apply, such as
  - Clustered sampling schemes
  - Growth curves and other random coefficient settings
  - Grouped experimental trials, e.g. multicenter clinical trials in medicine, or group randomized trials in education and social sciences

Computing

I anticipate that this course will depend strongly on three software packages. Each student is expected to be able to install and run this software on his or her own (laptop?) computer.

- `R`, a statistical analysis and programming environment (I am running version 3.3.1).
  See **http://cran.r-project.org/** We will install and use several libraries for R also. These include `arm`, `MASS`, `ggplot2`, `lattice`, and `r2jags`. See **http://cran.r-project.org/web/packages/**
- `jags`, a package for doing Bayesian statistics.
  See **http://mcmc-jags.sourceforge.net/**
- `rube`, an unpublished R library for using `jags` and WinBUGS from R.
  See **http://www.stat.cmu.edu/~hseltman/rube/**

We may briefly explore some other software (e.g. stan, http://mc-stan.org/), but the above are our main tools.

Homework, Tests and Grading

Your grade will be based on the following general categories:

Class participation	10%
Homework & Quizzes	40%
Midterm	15%
Final	15%
Project	20%

- *Homework:* graded generously. Serious-minded effort will be traded off against correctness.
- *Quizzes:* short, usually un-announced. They will be used as spot-checks for understanding, and graded on a generous pass-fail basis.

- *Midterm and Final:* in-class exams, held during and limited to regular class time.
- *Project:* In the past I have assigned two projects for this course (one at midterm and one at the end of the semester). This year I am going to assign just one project. An "interim report" will be due around midterm, and a "final report" will be due at the end of the semester.

  You do not have to find your own data/problem for the project. But if you have an idea for a project for this class (from your own work elsewhere), you may be able to do your project instead of the class project. *Please talk to me asap if you are interested in this.* I will consider written proposals until 5pm Friday September 23, and let you know in writing whether your project is approved. However, the sooner you talk with me about this, the more likely that we can craft something useful to both you and me.

Prepare your work with a word processor wherever you can. Be judicious about what computer material you include: include enough so we can give you credit for understanding & communicating what you are doing, but not so much that we have to search through pages to find the important parts (if so, we will skip your work and not give you credit).

Because of the large number of students in the class, late assignments (homework, take-homes) will not be accepted without explicit prior arrangements.

Academic Integrity

As members of a top-ranked academic institution, your academic integrity is assumed and expected. Here are some specific guidelines for this course:

Homework: You may work together on homework assignments; however you *must* list the names of the people you worked with, if you do so. If I ask you to change who you are working with, you must abide by my directions.

Quizzes, Exams and Project Work: You must do quizzes, exams and project work individually, without making use of other students' work or other unauthorized resources. Since project work is "take home", we rely on your honor.

All Work: If you get ideas or words from a website, journal article, book, another person (in or out of this class), etc., cite the source in your writeup, right where you use it. Then put a bibliography or list of sources cited at the end of the writeup.

General guidelines listed at `http://www.cmu.edu/policies/documents/Academic Integrity.htm`. However, I expect each of you to behave well above these lower bounds.

Disability and other Special Needs

Carnegie Mellon makes great efforts to provide physical and programmatic campus access to everyone. Disability Resources ensures that qualified individuals receive reasonable accommodations and that they further receive the rights and protections to equal access programs and services as guaranteed by the Americans With Disabilities Act (ADA) and Section 504 of the Rehabilitation Act of 1973.

If you have a documented disability, please let me know so that we can take whatever steps are needed to accomodate your needs.

Please contact CMU's Disability Resources office (http://www.cmu.edu/hr/eos/disability/) if

- You think you may have a disability and want to document it;
- You have a documented disability that is not being adequately accomodated.

For other issues and special needs, please contact me, your advisor or another trusted mentor, and/or the Office of the Dean of Student Affairs (http://www.studentaffairs.cmu.edu/dean/).

Tentative Schedule

The schedule below is somewhat tentative, especially in the later weeks of the semester. Please visit http://www.stat.cmu.edu/~brian/463-663 regularly for updates. G&H refers to the Gelman & Hill text.

Week	Date	Tentative Topics	Tentative Sources
Week 1	Aug 30, Sep 01	Course Introduction, Examples, R	G&H Ch 1, Handouts
Week 2	Sep 06, 08	Background, More on R	G&H Ch 2, Handouts
Week 3	Sep 13, 15	Linear Regression and lm()	G&H Ch's 3&4
Week 4	Sep 20*, 22	Logistic Regression and glm()	G&H Ch's 5&6
Week 5	Sep 27, 29*	Misc Topics: Simulation, Causal Reasoning**	G&H Ch's 7–9
Week 6	Oct 04, 06*	Multilevel Models and lmer()	G&H Ch's 11–13
Week 7	Oct 11, 13	Generalized mlm's, examples (PROJECT INTRODUCED)	G&H Ch's 14&15
Week 8	Oct 18, 20*	Oct 18: REVIEW Oct 20: IN-CLASS MIDTERM EXAM Oct 20: PROJECT INTERIM REPORT DUE	TBA
Week 9	Oct 25, 27	Misc Topics: Variable Selection, Methods of Estimation**	G&H Ch's 13, 18, Lynch Ch 2, & Class Notes
Week 10	Nov 01, 03	From ML to Bayes	Lynch Ch 3 and/or Handouts
Week 11	Nov 08, 10	Simulation-based Inference	Lynch, Ch's 4–6, Lecture Notes and/or Handouts
Week 12	Nov 15, 17	Using BUGS for Multilevel Models; Model Criticism	G&H Ch's 16–18, Lynch, Ch's 7–9, and/or Handouts
Week 13	Nov 22	Examples**	TBA
Week 14	Nov 29, Dec 01	Selected Topics: Analysis of Variance, Power, Missing Data . . . **	Selections from G&H Ch's 20–22, 25, or TBA
Week 15	Dec 06, 08	Dec 06: REVIEW & FINAL INSTRUCTIONS FOR PROJECT FINAL REPORT Dec 08: IN-CLASS FINAL EXAM	TBA
Finals Wk	Dec 13, 15	PROJECT FINAL REPORT DUE — DATE TBA	

* Dr. Junker may miss the 3pm office hours this day. Stay tuned for other arrangements.

** We will lengthen, shorten or change this topic depending on our needs that week.

The appendices of G&H contain brief, but *very* useful advice! If you refer to and follow the advice in appendices A and B, your work in all of the 36-4xx, 6xx and 7xx courses, and as an applied statistician, will be much better!