36-763: Hierarchical Linear Models  
Fall 2017 (Mini 2)  
TTh 3:00 - 4:20, BH 235A  
https://canvas.cmu.edu/ 

Course Information  

Instructor:  
Brian Junker, Statistics  
132E Baker Hall  
268-8874  
brian@stat.cmu.edu  

Office Hours (BH 132E):  
Immediately after class  
(or by appointment).  

TA:  
Neil Spencer  
FMS 328  
nspencer@andrew.cmu.edu  

Office Hours:  
232K Baker Hall, 3:00-4:00pm Mondays  
(or by appointment).  

Prerequisites  
There are no formal prerequisites for this class. However, I expect you to be familiar with statistical theory and the statistics of applied linear regression at a senior undergraduate or beginning graduate level.  
You will also be expected to know, or learn quickly, the computational software used for this course. That is, primarily R and stan and possibly jags (a variant of WinBUGS). It is also a good idea to know how to use the technical typesetting system L\LaTeX.  
Please feel free to contact me if you have any questions or need additional information.  

Required Text (Buy ASAP online)  

Other Worthwhile Texts  

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

Since the course is only 7–8 weeks long, we can do nothing in detail. The idea is to give you a taste of each topic, so that when you need to do this kind of analysis in the future, you will have some idea what details are important, and where to look to find out about them.

Computing

We’ll mostly be working in R with supporting libraries. You can (and should) install the relevant software to run on your own (laptop?) computer:

• R, a statistical analysis and programming environment (best to have the most current version).
  See http://cran.r-project.org/

• stan, a package for doing Bayesian statistics.
  See http://mc-stan.org/users/interfaces/rstan.html

• Various support libraries in R. These include arm, MASS, ggplot, lattice, plyer, etc.
  See http://cran.r-project.org/web/packages/

In previous semesters we used jags, a platform-neutral (runs on Mac, Windows, or Linux boxes) engine for traditional MCMC calculations, for fitting complex Bayesian models. This time, we will try to use stan, which employs Hamiltonian Monte Carlo for most model fitting. However, if there are things we can’t easily accomplish in stan, we will fall back to jags. If we use jags we will need the jags engine (http://mcmc-jags.sourceforge.net/), the r2jags R library (http://cran.r-project.org/web/packages/), and the rube R library (http://www.stat.cmu.edu/~hseltman/rube/).

We may digress into other software occasionally in the course, but the above will be our main tools.
Student Work

Homework:
I intend to give roughly 4–7 homework assignments, accounting for all of your grade. Homework will be a mix of developing and exploring theoretical material, and practicing using software to analyze data. The last homework assignment will also serve as a take-home final exam, and will be weighted to count roughly as much as two other homeworks.

You may work with other students on these problems or refer to other sources if you would like, unless otherwise indicated in my directions for the assignment. The computations and writeup of your assignment, however, must be your own. Please note that the written interpretation and conclusions from a data analysis are at least as important as generation of data summaries, statistics, tests, etc. If you work with others, or use any other sources, please list you collaborators and other sources on your assignment. See also the section on Academic Integrity below.

Prepare each hw as a single pdf: I find LaTeX very useful for this purpose, but MS-Word or even scanned handwriting (if it is legible) is fine. Your hw should be uploaded to the course’s canvas website.

Please label all output, plots, variables, etc., appropriately. Always be judicious about including computer output and graphs: show enough that we can clearly see what you are doing, but not so much that we will get lost or bored leafing through your work! A good rule of thumb is to remove any figures, tables, graphs, etc. that you do not have something interesting to say about in the text of your homework solution.

Special Projects:
Some students may request (or I may direct you) to do a special project in lieu of one or more homework assignments. Typically this would involve a data analysis & writeup for some data that you are interested in. If you think you have an idea for this, let me know ASAP so I can determine how & whether it will count as a special project for the class.

Because of the large number of students in the class, late assignments (homework, take-homes) will typically not be accepted.

Academic Integrity

As members of a top-ranked academic institution, your academic integrity is assumed and expected.

For 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.

Carnegie Mellon guidelines are listed at http://www.cmu.edu/academic-integrity/; 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 accommodate your needs.

1E.g., the last, “double” hw.
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 accommodated.

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 (https://www.cmu.edu/student-affairs/resources.html).

**Tentative Schedule**

Below is a tentative schedule of topics for the course. The readings are only approximate; specific readings will be assigned as we progress through the course. Please visit https://canvas.cmu.edu/ regularly for updates and current information. G&H refers to the Gelman & Hill (2007) text.

<table>
<thead>
<tr>
<th>Week</th>
<th>Date</th>
<th>Tentative Topics</th>
<th>Tentative Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 1</td>
<td>Oct 24, 26</td>
<td>Course Introduction, Preliminaries (you should review regression topics)</td>
<td>Email, class notes, handouts, G&amp;H Ch’s 1, 11 (G&amp;H Ch’s 3–6)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Week 2</td>
<td>Oct 31, Nov 2</td>
<td>Multilevel Models</td>
<td>G&amp;H Ch’s 12-15*</td>
</tr>
<tr>
<td>Week 3</td>
<td>Nov 7, 9**</td>
<td>Model Selection</td>
<td>Handouts &amp; class notes; G&amp;H Ch’s 16, 21</td>
</tr>
<tr>
<td>Week 4</td>
<td>Nov 14, 16</td>
<td>Bayesian Statistics &amp; MCMC</td>
<td>Handouts &amp; class notes; G&amp;H Ch 18</td>
</tr>
<tr>
<td>Week 5</td>
<td>Nov 21 (ONLY!)</td>
<td>Multilevel Models in STAN***</td>
<td>Handouts &amp; class notes; G&amp;H Ch’s 16-17</td>
</tr>
<tr>
<td>Week 6</td>
<td>Nov 28, 30</td>
<td>More STAN; Model Checking</td>
<td>G&amp;H Ch 24</td>
</tr>
<tr>
<td>Week 7</td>
<td>Dec 5, 7</td>
<td>Power, further reading</td>
<td>Handouts &amp; class notes; G&amp;H Ch 20</td>
</tr>
</tbody>
</table>

* These chapters contain the core material for the course, and will be referred to throughout the course.

** I will be out of town this day. Please stay tuned for guest lecturer, makeup lecture date, or other plans.

*** If STAN is not working out we will use JAGS instead.

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 as an applied statistician will be much better!