

Introduction to experimental design and model building for ecologists and evolutionary biologists Fall 2017

Class: W 10:30-12:30¹, HUH 142

Professor: Elizabeth Wolkovich (lizzie@oeb.harvard.edu)

Schedule at a glance²

Date	Topic	Leader(s)	Reading
6 Sep	Reproducibility ‘crisis’		in-class
13 Sep	Why learn statistics (and coding)?		pp. 0-56; Appendix B-C
20 Sep	Linear regression basics		pp. 59-102; Appendix A
27 Sep	Student project discussion		
4 Oct	Linear regression plus	Ben, Nick, Meghan	pp. 103-150
11 Oct	Generalized linear regression	Zane & Patrick	pp. 153-211
18 Oct	Design & sample sizes	Cat, Dave, Inbar	pp. 213-241
25 Oct	Understanding your model	Lydia, Dan, Ana	pp. 243-270
1 Nov	Causal inference		ARM: 1-6; 167-233
8 Nov	Catch-up on topics; pre-MM	Ian visits	
15 Nov	Multilevel models		ARM pp. 237-278 (rec: 279-299)
22 Nov	<i>Thanksgiving break</i>		
29 Nov	Final presentations!		

Course materials: There are two course textbooks. One is a new book by Andrew Gelman and Jennifer Hill called *Regression and Other Stories*, which covers linear regression and causal inference. You’ll receive a copy on the first day (which you cannot share with anyone outside of class, sadly). Any readings without a special note are from this book (though note that I may refer to this book as **ROS**). The other book you must track down a copy of (I suggest you buy one, they cost around \$60 and cover way more than we can in this course) *Data Analysis Using Regression and Multilevel/Hierarchical Models* by Gelman & Hill (more info here); note that I refer to this book as **ARM**. There are course reserves at the Cabot library for additional reading: *Statistical Rethinking* by MacElreath (more

¹Note that this class operates on standard time, not Harvard time.

²As this course is a seminar-style, its exact design will evolve somewhat throughout the term. Keep an eye out for an updated (and hyperlinked) syllabus on the Canvas site.

info here) and *The Statistical Sleuth* by Ramsey & Schafer (more info here).

What's due in each class?

Topic	Tasks (in addition to your book reading!)	Book problems
Reproducibility ‘crisis’	In-class reading	
Why learn statistics (and coding)?	Bring your laptop and vague project ideas to class! Video report 1, Video report 2; finish p-value readings	2.7: 1-4; 4.5: 4
Linear regression basics	Video report; survey	Example & ARM: 3.9: 4; Bonus: 7.10: 2
Student project discussion	Short talk on your possible project	
Linear regression plus	Stats in the news	8.10: 2; 9.9: 4
Generalized linear regression		10.10: 1
Design & sample sizes	Stats in the news	13.6: 3; 14.8: 2
Understanding your model		PPC problem (see Canvas)
Causal inference	Causal inference in the news	ARM: 9.10: 1
Catch-up on topics; review	Project check-in	
Multilevel models	Video report	
<i>Thanksgiving break</i>	Celebrate traditional American holiday in a manner of your choosing	
Final presentations!	Final project is due!	

Office hours: By appointment in HUH 223/Weld Hill (Arboretum).

Email: I check email once or twice a day between 1pm and 6pm (weekdays only). Please plan accordingly.

About the readings ... The class is centered around two textbooks and how well you can use what you learn from it and class to analyse data (more on that below). To get the most out of the class (and get a good participation score) you have to show me you have done the reading. This means re-reading as needed, reading additional materials when needed, doing assigned practice problems, going to see your classmates, then me, if you feel lost, and asking questions during class about what confused you. It also means answering questions in class, and asking questions to test your own knowledge. Your class participation makes up 30% of your grade, but you will get feedback from me in October and/or November if I have concerns about your performance.

Where do I find the data for the problems? You can find them via links from: <http://www.stat.columbia.edu/~gelman/arm/> (and yes, this should give you data for both books, but let me know if you have issues).

What is this whole leader thing? Four of the 12 class periods will have a team of student leaders. This means the week you are leader you will be in charge of highlighting the main points of the reading. You can do this however you like: ask critical questions, go over an example you have developed, have the class do an interesting project or related active-learning method to drive home the main messages. When possible I will assign you not just the usual reading but additional readings for background/additional information. Whatever you do, you will have up to **30 minutes of class time** and will be graded on: student engagement, how well you taught the take-home messages as well as how you defined them, and showing that all members of your group participated.

What is this whole video report thing? You have to watch a video and report on it in-class that week (just like the reading you should know what was covered and ask questions *in-class* as needed). The videos are generally hyperlinked to ‘Video report’ in the online PDF of the syllabus. I will also aim to post them on the Canvas site, but to repeat: the videos are generally hyperlinked to ‘Video report’ in the online PDF of the syllabus.

What is this whole stats in the news thing? Find an interesting use of statistics in the popular literature (e.g., *USA Today*, *New York Times*) and bring it to class with a plan to present it clearly in under two minutes to the class. During causal inference week you have to find a causal inference example.

Term data analysis project: One of the major goals of the class is for you to get your hands dirty analysing data (this includes: finding data, fitting models, interpreting your models and presenting your findings clearly). You can do this in three basic ways:

1. You have data from your MSc or PhD that you need to analyse. Now is your chance!
2. You don't have any data yet, but you want to design an experiment and will collect data from the literature to help you decide on your design (including your statistical design). This is a trickier option, but do-able.
3. You don't have any data yet, but you come up with a fun question for which you can gather data. For example, perhaps you would like to analyse the effect of climate change on weather during the Tour de France. You can outline the problem, gather the data and do an analysis on that.

You will have to outline your question, data, goal and basic plans for tackling the analysis in a document due at the start of class on 4 October. You will have the 27 September class period to discuss ideas and give feedback with your classmates. Your document and the feedback you provide to your classmates make up 25% of your grade. We'll check-in about progress towards your projects in class from 18 October - 8 November and you'll present your work (and submit a short written report) on 29 November, which counts for 35% of your grade.

Your final term project will include your code, simulated data, a short write-up with figures and a short presentation. More details to come in the official assignment.

When it is your week to present you project check-in, plan to give a 5-7 minute talk on your project (review the aim, the data, where you are at, present simulated fake data, what you learned from it and what you need help with). If you want to use slides you must send them to Lizzie by 7am the day of you talk in PDF format. This is ungraded but it is your major opportunity to get feedback on your project mid-term so doing poorly could translate in to doing poorly on the final submission of your project.

When am I leading? When am I doing my project check-in?*

Tentative

4 October: Linear regression plus (project updates on 18 October)

Nick nherrmann@g.harvard.edu

Ben benjaminrice@g.harvard.edu

Meghan blumstein@fas.harvard.edu

11 October: Generalized linear regression (project updates on 25 October)

Patrick pgorring@fas.harvard.edu

Jenny jpham@fas.harvard.edu

Zane rzwolf@g.harvard.edu

18 October: Design and samples size (project updates on 8 November)

Cat cchamberlain@g.harvard.edu

Dave davematthews@g.harvard.edu

Inbar imaayan@g.harvard.edu

25 Oct: Understanding your model (project updates on 1 November)

Dan dbuonaiuto@g.harvard.edu

Ana gonzalezvaldes@g.harvard.edu

Lydia lakrasilnikova@g.harvard.edu

Missing classes: You can miss up to one class without it impacting your grade. You cannot miss a class where you are a named ‘leader’ or where you are presenting. Note that we will cover a lot in each class and you are responsible for catching up on what you miss.

Grading:

In-class participation	30 points
Leading discussion	10 points
Your project overview and feedback to other students	25 points
Final presentation and project	35 points
Total	100 points