<u>EVE 298: Linear mixed modeling in ecology & evolution</u>

Taught by: Kate Laskowski (klaskowski@ucdavis.edu)

Dates: Winter or Spring quarter in odd years (last taught Spring 2021)

Grading: S/U

<u>Overview & goals:</u> This course will cover the statistical theory and practical application of general linear models and general linear mixed models using examples drawn from evolutionary and behavioral ecology. It will focus specifically on the analysis of clustered data, e.g. samples taken from the same lake, or measurements made on the same animal, which are common in ecological and evolutionary research. Students will learn 1) the assumptions of traditional linear models and 2) how such hierarchical or nested data violate these assumptions. Students will then learn 3) to build, select and validate mixed models that include random effects and 4) how to interpret and report such results for publication. The lecture portion of the course will focus on building the students' conceptual framework of the underlying statistics of such models and the discussion/lab section of the course will focus on the hands-on application of each week's knowledge for analyzing and interpreting example data sets in the statistical computing language R. The course is designed to build on knowledge students would have gained in a basic statistics (e.g. ANOVA, regression) or experimental design course and provide a firm conceptual foundation for students who wish to take more advanced statistical or modeling courses.

<u>Books & readings:</u> Many of the example datasets will come from textbooks such as Zuur et al.; however buying this book (or any of the other listed books) is not necessary for the course. All readings are considered supplementary to the course and to act as references in case there are topics you want more information on. The most appropriate papers (as I see them) are listed in the topics table below and all these pdfs are available in the Class Readings folder. The Files section has many more papers that you may find useful grouped by topic - have a look!

<u>Pre-requisites:</u> Basic statistics – This should *not* be the first statistics course you take. This course will assume knowledge of basic statistical concepts such as probability, means, variances, etc. Familiarity with t-tests, ANOVAs, regressions is necessary. Students will likely benefit most from this course in their 2-4th years.

Comfortable with R – students must be able to load packages, import data and have some understanding of basic coding in R. Registered students will be required to download and install R and several example data sets (links provided prior to course start).

Data sets – students who have datasets that may be appropriate for use in the course should prepare the datasets to be shared. This includes formatting them into a "tidy" format (each observation is a row; "long" format) and being able to explain the structure and motivation of the study. Note: not all datasets may be able to be analyzed; this will be determined at the discretion of the instructor.

<u>Lectures & labs:</u> All lectures and labs are held synchronously. I expect that students join and actively participate in these sessions, but recognize that life is crazy, especially now, so if there are circumstances that require you to miss lecture/lab, please contact me to work something out. These sessions will be recorded and available after the lecture (usually within a day or two).

To facilitate engagement and discussion, it helps to have videos enabled during class, but I realize there are situations or circumstances that make this difficult or impossible. Thus, I ask that have your camera on if you can, particularly during discussions and break-out rooms. I also encourage you to ask questions during lecture and discussion, which can be done either by using the 'raise hand' icon on Zoom or using the chat function.

Respect for privacy is an essential part of our classroom community. Images, text, screenshots, audio/video content from Zoom sessions may only be used for instructional purposes of this course. Participants should not distribute data captured from Zoom sessions to anyone outside the course, without appropriate consent from the individuals whose images/voice/data are involved. Unauthorized distribution or capture outside the course may violate federal or state privacy laws or University of California policies. This means, for instance, you should not post screenshots of your class, your instructor, or your classmates to social media without their express permission.

Final class project: A big component of this class is hands-on data analysis. The final project will be done in small groups by data provided by one of the group members. I will ultimately decide whether a dataset is appropriate for use in the final project and not all datasets may be chosen. Even if yours it not chosen, please know that you are still expected to participate fully in your group's analysis. This group work is consistently listed as one of the most valuable experiences by past students, even when they are analyzing someone else's data.

Class expectations and working together: Everyone is coming into this course from different fields and with different prior knowledge of and relationship towards statistics. The goal of this course is to give you a better foundation and confidence in doing basic (and not so basic) linear statistical modeling. This foundation and confidence can only come if you are honest about your own level of understanding. I would rather we discuss basic concepts in class then have folks feel left behind and get less out of the course than they otherwise could. So while I encourage you to work together, and in fact, you will on the final project, it is also important that homework be done by your own hands and reflect your own understanding.

Grading: S/U grading. A satisfactory grade requires in-class attendance and participation, completion of homework problems and full participation in the final group analysis project.

Sickness/Care responsibilities/absences: I understand that, especially during this global pandemic, everyone has unique situations and responsibilities which are constantly changing. My primary goal is to support you and your learning and I know that doing so sometimes requires flexibility. If you have illnesses, care responsibilities, or other scenarios that require you to be absent from class or are affecting your learning, please contact me so that I can work with you to develop a plan.

Even barring world events and personal situations, grad school can be a stressful time. If you are feeling especially stressed or just need to talk to someone, you should take advantage of the free counseling services offered on campus: https://shcs.ucdavis.edu/services/counseling-services. More links to mental health resources for graduate students can be found at: https://grad.ucdavis.edu/resources/help-and-support/mental-health-and-counseling . Finally, if you are living in Davis and you are having a hard time finding a healthy meal or getting basic necessities, please visit the UC Davis Pantry: https://thepantry.ucdavis.edu/ . Additional resources can also be found https://ebeler.faculty.ucdavis.edu/resources/faq-student-resources/

Topics: Below is the list of topics we will cover. They are listed in the order they will be presented; information on generalized linear mixed models will only be treated if there is time after the other topics are covered.

Topic	Goals	Readings
Linear model refresher	Identify components of linear model Understand difference between categorical versus continuous predictors Understand how degrees of freedom are used in a model R: How to interpret values in 'summary' (clam data)	Zuur Appendix A Zuur & Ieno 2016. Gelman & Hill Chapter 3 West et al. Chapter 2
Assumptions of linear models	Understand the four assumptions of linear model and Learn how to graphically check that these assumptions are met R: Graphical checks of model assumptions (clam, Loyn data)	Zuur Chapter 2 Gelman & Hill Chapter 3 Jacqmin-Gadda et al. 2007 Schielzeth et al. 2020
Linear mixed models	Identify non-independent structures in data Understand when and why we need to use random effects Understand how random effects differ from fixed effects conceptually and statistically R: Build mixed models & interpret summaries	Zuur Chapter 5 Sullivan et al. 1999. Dynamic Ecology "Is it fixed or random effect?" Harrison et al. 2018
Data exploration & model selection	Learn basic techniques to explore data graphically Learn how to build and select (if necessary) most appropriate models Understand pitfalls of different model selection techniques R: explore clam, Loyn, RIKZ data sets, perform model selection on RIKZ together, begging owls on your own	*Exploration:* Zuur Appendix A Zuur et al 2010 *Selection:* Zuur Chapter 5 Zuur & Ieno 2016 Schielzeth & Nakagawa 2013 Forstmeier & Schielzeth 2011 Barr et al. 2013 538.com graphic

Random effects and variances	Learn how to compare variances of random effects (repeatability) Learn how to identify levels of replication to identify the random structure of data (crossed versus nested) Understand what R-squared is and how to estimate it for a mixed models Learn about random slopes and when they are appropriate for a model R: R-squared in RIKZ dataset, complete Cetacean data set analysis	Nakagawa & Schielzeth 2010 Nakagawa & Schielzeth 2013 Harrison et al. 2018
Increasing interpretability of model outputs	Learn when and why standardizing variables is useful Learn about different centering techniques to increase interpretability of model outcomes	Schielzeth 2010. Nakagawa & Cuthill 2007 Gelman & Hill Chapter 4 Van de Pol & Wright 2009.
Other assumption violations	Learn when the assumption of variance homoegeneity is violated Learn how to relax this is assumption in linear models Understand what temporal autocorrelation is and why it is a problem in linear models Learn methods to help correct for this R: Deal with heterogeneity in squid data, temporal autocorr in moorhen data	Zuur Chapter 4, 6 Barrnett et al 2010 Shumway & Stoffer book
Generalized linear (mixed) models **(ONLY IF TIME AVAILABLE)**	Understand the properties of different distributions (Poisson and binomial) Understand when assuming these data distributions is useful for analysis Learn common problems and pitfalls with these types of models R: Deer disease prevalence data	Gelman & Hill Ch 5, 6, 14, 15 Bolker et al. 2009 Harrison 2014, 2015 Zuur Chapter 9, 10, 13 O'Hara & Kotze 2010

Books:

Zuur et al. "Mixed effects models and extensions in ecology with R". Springer Press

Gelman & Hill. "Data analysis using regression and multilevel/hierarchical models." Cambridge University Press

West, Welch & Galecki. "Linear mixed models: A practical guide using statistical software." CRC Press

Shumway & Stoffer. "Time series analysis and its applications." Springer press. FREE pdf: https://www.stat.pitt.edu/stoffer/tsa4/tsa4.pdf

Articles:

Barnett et al. 2010. Using information criteria to select the correct variance-covariance structure for longitudinal data in ecology. *Methods in Ecology & Evolution 1.*

Barr, D.J., Levy, R., Scheepers, C., Tily, H.J., 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. Journal of Memory and Language 68, 255–278. https://doi.org/10.1016/j.jml.2012.11.001

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., White, J.-S.S., 2009. Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology & Evolution 24, 127–135. https://doi.org/10.1016/j.tree.2008.10.008

Andrew Gelman post on standardizing: http://www.stat.columbia.edu/~gelman/research/unpublished/standardizing.pdf

Forstmeier & Schielzeth. 2011. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse. *Behav Ecol Sociobiol 65*

Harrison, X.A., 2015. A comparison of observation-level random effect and Beta-Binomial models for modelling overdispersion in Binomial data in ecology & evolution. PeerJ 3, e1114. https://doi.org/10.7717/peerj.1114

Harrison, X.A., 2014. Using observation-level random effects to model overdispersion in count data in ecology and evolution. PeerJ 2, e616. https://doi.org/10.7717/peerj.616

Harrison, X.A., Donaldson, L., Correa-Cano, M.E., Evans, J., Fisher, D.N., Goodwin, C.E.D., Robinson, B.S., Hodgson, D.J., Inger, R., 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6, e4794. https://doi.org/10.7717/peerj.4794

Jacqmin-Gadda et al. 2007. Robustness of the linear mixed model to misspecified error distribution. *Comp Stats & Data Analysis* 51

Nakagawa, S., Schielzeth, H., 2010. Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists. Biological Reviews 85, 935–956

Nakagawa & Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods Ecol & Evol 4

Nakagawa & Cuthill. 2007. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol. Reviews 82*

O'Hara & Kotze. 2010. Do not log transform count data. *Methods Ecol Evol 1.*

Sullivan et al. 1999. Tutorial in biostatistics: an introduction to hierarchical linear modeling. *Statistics in Medicine 18*

Schielzeth 2010. Simple means to improve the interpretability of regression coefficients. *Methods Ecol Evol*

Schielzeth & Nakagawa. 2013. Nested by design: model fitting and interpretation in a mixed model era. *Methods Ecol Evol 4.*

Schielzeth, H., Dingemanse, N.J., Nakagawa, S., Westneat, D.F., Allegue, H., Teplitsky, C., Réale, D., Dochtermann, N.A., Garamszegi, L.Z., Araya-Ajoy, Y.G., 2020. Robustness of linear mixed-effects models to violations of distributional assumptions. Methods in Ecology and Evolution 11, 1141–1152.

Van de Pol & Wright. 2009. A simple method for distinguishing within- and between-subject effects using mixed models. Anim Behav 77

Zuur et al. 2010. A protocol for data exploration to avoid common statistical problems. *Methods in Ecology & Evolution* 1, 3-14

Zuur & Ieno. 2016. A protocol for conducting and presenting results of regression-type analyses. *Methods Ecol Evol* 7.

538 graphic: https://fivethirtyeight.com/features/science-isnt-broken/#part1

Other useful resources:

- http://ase.tufts.edu/gsc/gradresources/guidetomixedmodelsinr/mixed%20model%20guide.html
- https://ourcodingclub.github.io/ Seriously helpful tutorials on lots of stats/R stuff (written by ecologists!)
- http://m-clark.github.io/documents.html More in depth tutorials, also more advanced topics (Bayesian, SEM, Generalized)
- http://www.bioinfo.org.cn/~wangchao/maa/w.statistic.pdf Wasserman's 'All of Statistics' book available as a pdf online