<u>EVE 225: Linear mixed modeling in ecology & evolution</u>

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

Meeting times:

- Lecture: Tu/Thurs 1:40pm – 3:00pm in Wellman 109
- Lab: Wednes 9:00am – 9:50am in TLC 3211
- Office hours: Mondays 4:00 – 5:00pm in Storer 2206

<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> 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).

Basic statistics – this course will assume knowledge of basic statistical concepts such as means, variances, covariates, etc. A basic understanding of t-tests, ANOVAs, regressions is strongly encouraged.

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>Attendance policy:</u> All lectures and labs will be held in person. There is a high demand for this course, with a long wait list s**o in order to guarantee your spot, you must attend the first lecture and lab meetings (January 10 & 11)!** After this, I will not be taking attendance, but please be aware that you are expected to make every effort to attend in person during designated 'group work' days. If you need to be absent for these days, you'll need to make arrangements with your group members. Participation in group work will be evaluated through the group member rating that each member will submit with their final report. Group member ratings are worth 15% of the total grade so if someone is rated as not participating and contributing as expected, this can drop your score by an entire letter grade.

Final projects: 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.

The goal of the final project is to give you experience conducting mixed model analyses, communicating your methods and results and then critiquing the analyses of others. We will build up to the final project by first conducting guided analyses in lab. Then you will perform an independent analysis with your group and write this up including the motivating questions, methods and results. This first draft will go through a round of 'peer review' with another group. Rubrics will be provided in class for the expectations of what should be included in the report. After this review, you will have an opportunity to revise your final report before submission to me. Finally, you will also provide a 'rating' for your fellow group members; this is done to help ensure equal commitment and contribution by each member of the group (i.e. no one does the lion's share and no one gets to freeload). Rubrics will be provided in class for expectations for each draft, the peer review process and the group member rating.

Grading: Letter grading.

Component	**Due date**	**Amount of grade**
'Practice' analysis report	Emailed to Kate 11:59pm Feb 17	20%
Independent analysis		
- Initial draft	Printed & brought to class March 9	25%
- Final draft	Emailed to Kate 11:59pm March 21	40%
- Group member rating	(submitted with final draft)	15%

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.

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 we 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/

Overall course learning objectives:

Students will be able to:

- Identify each parameter in a linear (mixed) and explain what it is estimating including the difference between ‘fixed’ and ‘random’ effects
- Translate between statistical models, experimental designs/research questions and graphs of results. (Your model = your question = your graph)
- Describe the assumptions of linear (mixed) models, how and when they are violated and what can be done to correct such violations
- Perform biologically-informed model selection analyses
- Clearly present and describe the methods used in analyses and the results in written format similar to what would be done in a scientific manuscript
- Perform constructive peer review of the methods and results of other researchers
- *Use google more effectively to find the answers to statistical and coding problems*

Dates	Topics	Further readings
Week 1 Jan 10 – 12 Linear model parameters	Conceptual and statistical difference between categorical versus continuous predictors Translating between research questions, statistical models and results' graphs **Lab 1:** Identify and explain how to interpret values in 'summary' (clam data)	Zuur Appendix A Zuur & Ieno 2016. Gelman & Hill Chapter 3 West et al. Chapter 2
Week 2 Jan 17 – 19 model assumptions	Assumptions of linear model and when/how they are violated Relationship between random effects, degrees of freedom and replication **Lab 2:** graphically check assumptions (clam, Loyn data); data exploration	Zuur Chapter 2 Gelman & Hill Chapter 3 Jacqmin-Gadda et al. 2007 Schielzeth et al. 2020 Zuur Appendix A Zuur et al 2010
Week 3 Jan 24 – 26 Introduction to random effects	Conceptual and statistical difference between 'random' effects and 'fixed' effects **Lab 3:** model selection with RIKZ data	Zuur Chapter 5 Sullivan et al. 1999. Dynamic Ecology "Is it fixed or random effect?" Harrison et al. 2018
Week 4 Jan 31 – Feb 2 Model selection	Model selection philosophies Types of research and how they differ statistically **Lab 4:** 'Practice' analysis with owl dataset ***By Feb 2: Describe datasets in google sheet!***	Harrison et al. 2018 Zuur Chapter 5 Zuur & Ieno 2016 Jaeger & Halliday 1998 Schielzeth & Nakagawa 2013 Forstmeier & Schielzeth 2011 Barr et al. 2013 Whittingham et al 2006
Week 5 Feb 7 – 9 model interpretation	Model interpretation including interaction effects and building results' tables Familiarize yourselves with group work research questions **Lab 5:** 'Practice' analysis with owl dataset ***By Feb 9: Assign yourselves into groups of 3-4 on selected datasets***	Harrison et al. 2018 Zuur & Ieno. 2016 Schielzeth & Nakagawa 2013
Week 6 Feb 14 – 16 Multiple variance components	Multiple variance components including nested and crossed effects Estimating proportion of variance explained (repeatability and R-squared for mixed models)	Nakagawa & Schielzeth 2010 Nakagawa & Schielzeth 2013

	Lab 6: R-squared ***Owl analysis report due 11:59pm Feb 17th!***	
Week 7 Feb 21 – 23 Increasing interpretability of model outputs	Standardizing predictors to ease interpretation and comparison Centering (including within-subject) of predictors **Lab 7:** model selection with cetacean dataset	Schielzeth 2010 Nakagawa & Cuthill 2007 Gelman & Hill Chapter 4 Van de Pol & Wright 2009.
Week 8 Feb 28 – Mar 2 Random regression	Random regression models Temporal auto-correlation **Lab 8:** Group work – build analysis plan	Barnett et al. 2010 Martin et al. 2011
Week 9 Mar 7 – 9 Independent analysis	Group work analysis together **Lab 9:** group work analysis ***Analysis drafts must be printed and brought to class on Mar 9th***	
Week 10 Mar 14 – 16 Peer review	(Tues) Group work: provide peer review to each other's groups ***Give Kate copy of peer review (written on draft report is okay)*** (Thurs) Group work: Any final group analysis details! **Lab 10:** Group work analysis together	
Final	***Final draft to Kate 11:59pm on March 21***	

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

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., 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

Martin, J.G., Nussey, D.H., Wilson, A.J., Réale, D., 2011. Measuring individual differences in reaction norms in field and experimental studies: a power analysis of random regression models. Methods in Ecology and Evolution 2, 362–374.

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*

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

Extra reading on other topics you might be interested in:

Araya-Ajoy, Y.G., Mathot, K.J., Dingemanse, N.J., 2015. An approach to estimate short-term, long-term and reaction norm repeatability. Methods in Ecology and Evolution 6, 1462–1473. https://doi.org/10.1111/2041-210X.12430

Arnold, T.W., 2010. Uninformative Parameters and Model Selection Using Akaike's Information Criterion. The Journal of Wildlife Management 74, 1175–1178. https://doi.org/10.1111/j.1937-2817.2010.tb01236.x

Banner, K.M., Irvine, K.M., Rodhouse, T., 2020. The Use of Bayesian Priors in Ecology: The Good, The Bad, and The Not Great. Methods in Ecology and Evolution n/a. https://doi.org/10.1111/2041-210X.13407

Björklund, M., 2019. Be careful with your principal components. Evolution 73, 2151–2158. https://doi.org/10.1111/evo.13835

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

Brommer, J.E., Kontiainen, P., Pietiäinen, H., 2012. Selection on plasticity of seasonal life-history traits using random regression mixed model analysis. Ecology and Evolution 2, 695–704. https://doi.org/10.1002/ece3.60

Dingemanse, N.J., Dochtermann, N.A., 2013. Quantifying individual variation in behaviour: mixed-effect modelling approaches. Journal of Animal Ecology 82, 39–54. https://doi.org/10.1111/1365-2656.12013

Gomes, D.G.E., 2022. Should I use fixed effects or random effects when I have fewer than five levels of a grouping factor in a mixed-effects model? PeerJ 10, e12794. https://doi.org/10.7717/peerj.12794

Grueber, C.E., Nakagawa, S., Laws, R.J., Jamieson, I.G., 2011. Multimodel inference in ecology and evolution: challenges and solutions. Journal of Evolutionary Biology 24, 699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x

Hadfield, J.D., Wilson, A.J., Garant, D., Sheldon, B.C., Kruuk, L.E.B., 2010. The Misuse of BLUP in Ecology and Evolution. The American Naturalist 175, 116–125. https://doi.org/10.1086/648604

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

Jaeger, R.G., Halliday, T.R., 1998. On confirmatory versus exploratory research. Herpetologica 54, S64–S66.

Mitchell, D.J., Dujon, A.M., Beckmann, C., Biro, P.A., 2020 Temporal autocorrelation: a neglected factor in the study of behavioral repeatability and plasticity. Behav Ecol. https://doi.org/10.1093/beheco/arz180

Morrissey, M.B.; R., 2018. Multiple Regression Is Not Multiple Regressions: The Meaning of Multiple Regression and the Non-Problem of Collinearity. Philosophy, Theory, and Practice in Biology 10. https://doi.org/10.3998/ptpbio.16039257.0010.003

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

Whittingham, M.J., Stephens, P.A., Bradbury, R.B., Freckleton, R.P., 2006. Why do we still use stepwise modelling in ecology and behaviour? Journal of Animal Ecology 75, 1182–1189. https://doi.org/10.1111/j.1365-2656.2006.01141.x