

# [ECOLOG-L] ONLINE COURSE " Model selection and model simplification

ECOLOG-L@community.esa.org on behalf of

Oliver Hooker <privateemail54982@community.esa.org>

Tue 4/13/2021 6:24 AM

To: ECOLOG-L@community.esa.org <ECOLOG-L@community.esa.org>;

ONLINE COURSE " Model selection and model simplification (MSMS01) This course will be delivered live

<https://www.prstatistics.com/course/model-selection-and-model-simplification-msms01/>

TIME ZONE " UK local time (GMT+0) " however all sessions will be recorded and made available allowing attendees from different time zones to follow a day behind with an additional 1/2 days support after the official course finish date (please email [oliverhooker@prstatistics.com](mailto:oliverhooker@prstatistics.com) for full details or to discuss how we can accommodate you).

## Course Overview:

This two day course covers the important and general topics of statistical model building, model evaluation, model selection, model comparison, model simplification, and model averaging. These topics are vitally important to almost every type of statistical analysis, yet these topics are often poorly or incompletely understood. We begin by considering the fundamental issue of how to measure model fit and a model's predictive performance, and discuss a wide range of other major model fit measurement concepts like likelihood, log likelihood, deviance, residual sums of squares etc. We then turn to nested model comparison, particularly in general and generalized linear models, and their mixed effects counterparts. We then consider the key concept of out-of-sample predictive performance, and discuss over-fitting or how excellent fits to the observed data can lead to very poor generalization performance. As part of this discussion of out-of-sample generalization, we introduce leave-one-out cross-validation and Akaike Information Criterion (AIC). We then cover general concepts and methods related to variable selection, including stepwise regression, ridge regression, Lasso, and elastic nets. Following this, we turn to model averaging, which is an arguably always preferable alternative to model selection. Finally, we cover Bayesian methods of model comparison. Here, we describe how Bayesian methods allow us to easily compare completely distinct statistical models using a common metric. We also describe how Bayesian methods allow us to fit all the candidate models of potential interest, including cases where traditional methods fail.

THIS IS ONE COURSE IN OUR R SERIES " LOOK OUT FOR COURSES WITH THE SAME COURSE IMAGE TO FIND MORE IN THIS SERIES

Email [oliverhooker@prstatistics.com](mailto:oliverhooker@prstatistics.com) with any questions

Wednesday 14th " Classes from 12:00 to 20:00

Topic 1: Measuring model fit. In order to introduce the general topic of model evaluation, selection, comparison, etc., it is necessary to understand the fundamental issue of how we measure model fit. Here, the concept of conditional probability of the observed data, or of future data, is of vital importance. This is intimately related, though distinct, to concept of likelihood and the likelihood function, which is in turn related to the concept of the log likelihood or deviance of a model. Here, we also show how these concepts are related to concepts of residual sums of squares, root mean square error (rmse), and deviance residuals.

Topic 2: Nested model comparison. In this section, we cover how to do nested model comparison in general linear models, generalized linear models, and their mixed effects (multilevel) counterparts. First, we precisely define what is meant by a nested model. Then we show how nested model comparison can be accomplished in general linear models with F tests, which we will also discuss in relation to  $R^2$  and adjusted  $R^2$ . In generalized linear models, and mixed effects models, we can accomplish nested model comparison using deviance based chi-square tests via Wilks's theorem.

Topic 3: Out of sample predictive performance: cross validation and information criteria. In the previous sections, the focus was largely on how well a model fits or predicts the observed data. For reasons that will be discussed in this section, related to the concept of overfitting, this can be a misleading and possibly even meaningless means of model evaluation. Here, we describe how to measure out of sample predictive performance, which measures how well a model can generalize to new data. This is arguably the gold-standard for evaluating any statistical models. A practical means to measure out of sample predictive performance is cross-validation, especially leave-one-out cross-validation. Leave-one-out cross-validation can, in relatively simple models, be approximated by Akaike Information Criterion (AIC), which can be exceptionally simple to calculate. We will discuss how to interpret AIC values, and describe other related information criteria, some of which will be used in more detail in later sections.

Day 2

Thursday 15th "Classes from 12:00 to 20:00

Topic 4: Variable selection. Variable selection is a type of nested model comparison. It is also one of the most widely used model selection methods, and variable selection of some kind is almost always done routinely in all data analysis. Although we will also have discussed variable selection as part of Topic 2 above, we discuss the topic in more detail here. In particular, we cover stepwise regression (and its limitations), all subsets methods, ridge regression, Lasso, and elastic nets.

Topic 5: Model averaging. Rather than selecting one model from a set of candidates, it is arguably always better perform model averaging, using all the candidates models, weighted by the predictive performance. We show how to perform model average using information criteria.

Topic 6: Bayesian model comparison methods. Bayesian methods afford much greater flexibility and extensibility for model building than traditional methods. They also allow us to easily directly compare completely unrelated statistical models of the same data using information criteria such as WAIC and LOOIC. Here, we will also discuss how Bayesian methods allow us to fit all models of potential interest to us, including cases where model fitting is computationally intractable using traditional methods (e.g., where optimization convergence fails). This allows us therefore to consider all models of potential interest, rather than just focusing on a limited subset where the traditional fitting algorithms succeed.

~~~~~

Manage your Group settings:

- \* To reply to sender (Oliver Hooker):  
privateemail54982@community.esa.org
- \* To view sender's profile, click this link:  
<http://community.esa.org/profile.htm?mode=vp&pid=54982>
- \* To post to entire group, send email to:  
ECOLOG-L@community.esa.org
- \* To email the group moderator(s) (George, Jonathan, Maya Sohn, Hattie):  
ECOLOG-L-moderator@community.esa.org
- \* To UNSUBSCRIBE or change delivery settings, click this link:  
<http://community.esa.org/p.htm?m=qs&em=david.schneider@mun.ca&sk=pvnB7Y>
- \* To UNSUBSCRIBE from this group, send email to:  
ECOLOG-L-unsubscribe@community.esa.org
- \* To remove yourself from this group, send email to:  
ECOLOG-L-remove@community.esa.org
- \* To visit this group on the web, click this link:  
<http://community.esa.org/group.htm?igid=35455>

Powered by GoLightly, <http://www.golightly.com/>

-----