FISH 6003: Statistics and Study Design for Fisheries Science

Dr. Brett Favaro

Winter 2018

E-mail: brett.favaro@mi.mun.ca Web: mifisheriesscience.github.io/courses/6003Stats/
Office Hours: Thursdays 0900-1200 Class Hours: M 9-950, T 9-1050
Office: Marine Institute W2009 Class Room: M W2041, T W1003

Deriving trends from data is a key aspect of fisheries science. In this course, students will gain expertise in some of the most powerful (and common) statistical techniques. Some concepts, such as generalized linear models, will be covered in extensive detail as they are relevant throughout the life sciences. Other topics will be more specific to fisheries. Equal emphasis will be placed on designing powerful studies and in using pre-existing data to answer scientific questions.

The course will be taught from a user perspective - while theory will be covered as needed, emphasis will be placed on when, where, and how to use these models.

Learning Outcomes

By the end of this course, students will develop a working competency of R Statistical Software, and will be introduced to the software environment that fisheries scientists must master. Great importance will be placed on fostering the ability to self-teach and stay abreast of developments in data collection, management, display, and programming.

The fundmental competencies that you will develop in this course are:

- Able to design a statistically powerful study
- Able to define, build, and run an appropriate model for a dataset, using R Statistical Software
- Understand regression-type analysis ranging from simple (single linear regression) to complex (generalized linear mixed effects models) and how to apply them
- Understand and be able to use model selection approaches to identify the
- Use power analysis to determine the size of study needed
- Understand the difference between Bayesian and frequentist methods, and when each might be used
- Be able to apply the technique of meta-analysis to measure weight of evidence on a given topic

We will focus extensively on regression-type models due to their ubiquity in the life sciences, and their ability to be applied to many relevant situations within fisheries.

Expectations and Aspirations

Statistics can be intimidating, but that doesn't need to be the case. This class will dymystify common statistical techniques that are relevant to pretty much anyone in fisheries science, and will help you learn how to do them yourself in R. We will start with the fundamentals and conclude with some fairly advanced analysis. We will focus on statistical principles, such that you are empowered to self-study and learn new techniques after the class is over.

My expectations are that you engage meaningfully with the course material. Come in and do your best, and don't be afraid to make mistakes. This will be a supportive environment

Course Structure

The course will meet twice weekly - one 1-hr block and one 2-hr block. Lectures will balance theory and practice, and there will be in-class activities, primarily on the Tuesday class.

Generally, a topic will be introduced within a traditionally designed lecture on Monday. On Tuesday, we will discuss how to implement the theory into practice, using R. In-class activities will tend to fall on Tuesdays.

Reference Books

We will mostly rely on primary literature in this course. I do not require that you buy any books for this course, although I have compiled a list of some key useful references below.

Important papers

Data exploration: Zuur, A.F., Ieno, E.N., and Elphick, C.S. (2009). A protocol for data exploration to avoid common statistical problems. Methods in Ecology and Evolution 1:1, 3-14.

LMs and GLMs: Introudction to Generalized Linear Models. STAT 504: Analysis of Discrete Data, Penn State Eberly College of Science

Mixed models:

- Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., and White, J-S.S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. Trends in Ecology and Evolution 24:3, 127-135. See also: http://glmm.wikidot.com/
- Nakagawa, S., and Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. Methods in Ecology and Evolution 4, 133-142.
- GLMM worked examples, by Dr. Ben Bolker

Power analysis: Kain, M.P., Bolker, B.M., and McCoy, M.W. (2015). A practical guide and power analysis for GLMMs: detecting among treatment variation in random effects. PeerJ e1226. doi: 10.7717/peerj.1226

Conducting and reporting regressions: Zuur, A.F., and Ieno, E.N. (2016). A protocol for conducting and presenting results of regression-type analyses. Methods in Ecology and Evolution 7:5, 636-645.

Meta-analysis: Harrison, F. (2011). Getting started with meta-analysis. Methods in Ecology and Evolution 2:1, 1-10.

Useful online courses

- GLMM course on Github, By Dr. Sean Anderson
- Biology 501: Quantitative Methods in Ecology and Evolution, by Dr. Dolph Schluter
- STAT 545: Data wrangling, exploration, and analysis with R, by Dr. Jenny Bryan

Useful textbooks

- Koricheva, J., Gurevitch, J., and Mengersen, K. (2013). Handbook of meta-analysis in ecology and evolution. Princeton University Press, 520 pp.
- Zuur, A., Ieno, E.N., Walker, N., Saveliev, A.A., and Smith, G.M. (2009). Mixed effects models and extensions in ecology with R. Springer, 574 pp.

Course Policies

Social Media

Students are welcome to tweet about the course using the hashtag #MIStats - but the Chatham House Rule is in effect. That means you **may not reveal the identity of the person speaking** in your tweets without their express permission. We want to encourage people to actively participate and make mistakes without fear of their mishaps being broadcast across the world.

Code of Conduct

You have the right to expect a supportive, safe environment in this course. This course will be governed by my Fisheries Science Code of Conduct, which all participants are expected to respect.

Digital Competency

Students are expected to have basic computer competency. You should be able to operate Microsoft Word, Powerpoint, and Excel, or equivalent (e.g. OpenOffice or Google Docs). You should be able to download and install software onto your computer. Please install R Statistical Software and RStudio prior to begining the course.

If you lack these skills, please consult training materials on your own time. **Please bring a laptop to every class.**

Students should be proficient with R Statistical Software. Normally, students will have completed FISH 6002 or equivalent training prior to starting FISH 6003. FISH 6002 is not a formal pre-requisite - you may demonstrate R competency in other ways.

E-mail Policy

E-mail is not a primary tool for communication in this class. If you have questions about course content, your order of operation should be:

1. Check the syllabus
2. Ask in class, or discuss with colleagues
3. Ask on Slack (this way, everyone can benefit from an answer)
4. Request a meeting with me (normally, to be held during office hours)

If emailing me a meeting request, use the subject line "FISH 6003: Meeting request." Please indicate three potential meeting times (Start with my office hours. Only if those don't work, propose alternatives) and explain in 1-3 lines what you want to meet about.

E-mail is impersonal, burdeonsome, and adds to confusion, so let's minimize it.

Class Participation

There will be a LOT going on in this class. Most assignments are designed to be completed mostly in-class time. The class is highly collaborative, meaning you need to be present to do it.

Accomodations will be made for serious illness or other extenuating circumstances. However, it is the student's responsibility to stay caught up with course materials - and missing in-class activities will result in a decreased participation grade.

So please, don't make it your plan to miss class!

Academic Honesty

This course is governed by MUN's regulations on academic misconduct.

Course Schedule

Week	Dates	Theme
1*	*January 8 and 9*	[Introduction, and the Philosophy of Statistics]
2	January 15 and 16	[Data Exploration]
3	January 22 and 23	[Simple Linear Regression]
4	January 29 and 30	[Multiple Linear Regression]
5	February 5 and 6	[Model Selection]
6	February 12 and 13	[Generalized Linear Models]
X	February 19-23	[*Winter semester break*]
7	February 26 and 27	[Mixed Models]
8	March 5 and 6	[Generalized Linear Mixed Models]
9	March 12 and 13	[Power Analysis]
10	March 19 and 20	[Intro to Bayesian Statistics 1]
11	March 26 and 27	[Intro to Bayesian Statistics 2]
12	April 3	[Meta-Analysis]

- I will be absent on the first week of courses. Over the first few weeks, we will schedule three hours that work for everyone to make up for the missing class time. **Our first in-class lecture will be Jan 15**

Assignments and Grading

10% of your course grade is earned by participation. Just show up, be yourself, and participate!

The remainder will be earned by completing assignments. These fall into three categories:

- Presentation on a special topic (20%)
- Major assignment (50%)
- Minor assignments (20%)