Stat 547N/FISH 506H - Statistics in Ecology and Marine Sciences

Syllabus

Tuesday/Thursday 15:00-16:30 pm, Sep 4th – Oct 20th 2022

Instructor: Marie Auger-Méthé, auger-methe@stat.ubc.ca
Office hours details found on Canvas

Most information and assignments accessible via Canvas

Data in ecology and marine sciences are frequently associated with large challenges. Controlled experiments are often difficult and observational studies are often associated with missing data and measurement error. This class will introduce some of the challenges of using statistics to answer questions in ecology and marine sciences and the statistical tools developed to handle them. Topics covered in this class are: missing data, multiple imputation, censored and truncated data, generalised linear models (GLMs), overdispersion, hidden Markov models, and state-space models.

Prerequisites: *Must be eligible to enrol in a graduate class. This course is a statistics class for graduate students in the Department of Statistics (STAT) and the Ocean and Fisheries Graduate program (OCF), but interested students from other departments are welcome to enrol via FISH 506H. The course is intended for students with a good statistics background and some familiarity with R. It is most suitable for MSc and PhD students in Statistics and PhD students in other scientific fields.*

Tentative schedule:

Date	Lecture	Paper discussion	R tutorial	Assignment Due
Sep 2 -*Tu*	*No class –* ***Imagine UBC*** *https://students.ubc.ca/new-to-ubc/orientations/imagine-ubc*			
Sep 4 –*Th*	Intro to class + challenges of ecological data + review of statistical approaches	--	--	--
Sep 9 -*Tu*	Missing data	p-value	--	Paper summary & questions - *submit by noon*
Sep 11 -*Th*	--	--	Missing data	--
Sep 16 -*Tu*	Truncated + Setting groups	--	--	--
Sep 18 -*Th*	Censored data	--	--	--
Sep 23 -*Tu*	--	--	Censored data	--
Sep 25 -*Th*	GLMs + Overdispersion	AIC	--	Paper summary & questions - *submit by noon*

Sep 30 -*Tu*	*No class – **National Day for Truth & Reconciliation*** *https://orangeshirtday.org* *https://www.canada.ca/en/canadian-heritage/campaigns/national-day-truth-reconciliation.html*			
Oct 2 -*Th*	--	--	GLMs	Final project outline
Oct 7 -*Tu*	Mixture models	--	--	Formative peer evaluations
Oct 9 -*Th*	Hidden Markov models	--	HMM - *short & not graded*	--
Oct 14 -*Tu*	State-space models	--	SSM - *short & not graded*	--
Oct 16 -*Th*	Project presentations	--	--	--
Oct 24 -*Fr*	(No class)			Project report + summative peer evaluations

Auditing: *students interested in auditing must do so formally and must attend all classes and tutorials. Auditing student attendance is required as the course has many group-based and interactive components, and auditing students will be assigned to a group. Auditing students will be asked to do the readings before the in-class discussion, but will not be asked to complete any of the assignments.*

Assessments:

1. Paper discussions: 25% total (12.5% each)
2. R tutorials: 30% total (10% each)
3. Group final project: 41% total (5% outline, 16% presentation, and 20% written report)
4. Group participation: 4%

1. Paper discussion guidelines and associated assignments:

The course will have 2 discussion sessions. The class will be assigned scientific papers to read in advance and will be asked to hand in an assignment before the in-class discussion. Each student will be asked to summarise *in their own words* the papers assigned that week. The summary of the papers should be only one paragraph and no longer than 300 words. If multiple papers were assigned for a session, make sure to highlight whether they have contrasting views on the topic. In addition, each student will be asked to write down 2 discussion questions that arose from reading the papers. Find questions that should spark debate and that are specific to the topic in question. The questions should be inspired/informed by the papers. Please, submit a pdf file on Canvas before noon the day of the class discussion. *You cannot use genAI tools such as ChatGPT to help you write this assignment; see details below.*

Topic	**Paper**	**Link** ***see below for how to access off campus****
p-value		
	Wasserstein & Lazar (2016). The ASA's statement on p-values: context, process, and	http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1154

	purpose. The American Statistician 70:129-133	108?needAccess=true
	Ionides *et al.* (2017). Response to the ASA's statement on p-values: context, process, and purpose. The American Statistician 71:88-89	http://amstat.tandfonline.com/doi/pdf/10.1080/00031305.2016.1234977?needAccess=true
	Dushoff *et al.* (2019). I can see clearly now: reinterpreting statistical significance. Methods in Ecology and Evolution 10:756-759	https://besjournals.onlinelibrary.wiley.com/doi/full/10.1111/2041-210X.13159
AIC		
	Sutherland *et al.* (2023). Practical advice on variable selection and reporting using Akaike information criterion. Proceedings of the Royal Society B 290:20231261	https://royalsocietypublishing.org/doi/epdf/10.1098/rspb.2023.1261

* Need to either be on campus to access the link or use the EZ-proxy tools from the library if off-campus (https://services.library.ubc.ca/electronic-access/connect/ezproxy-toolkit/)

2. R tutorials:
These tutorials will be described in a separate document available as an assignment on *Canvas* on the appropriate day. Make sure you have R and R Studio installed on your laptop before the first tutorial. I am expecting that you are familiar with R. These tutorials are intended to be completed in class. I will be in class to answer any questions you may have. *You cannot use genAI tools such as ChatGPT to help you complete these tutorials; see details below.*

3. Final project:
The goal of the final project is to explore some of the analyses covered in class in more details, and, in particular, to learn to apply these analyses to real data. The project should focus on exploring how different methods perform. You can either use your own data or free data available online. This will be a group project and I am open to various project ideas as long as you explore some of the methods covered in class (e.g., GLMs, Multiple Imputation) and show how different ways to analyse the data affect the results. While it is important that at least one of the analyses of the project is one of the methods covered in class, you are welcome to explore other techniques. *You cannot use genAI tools such as ChatGPT to help you complete this project; see details below.*

You will be assigned to a group in the fourth class. There will be a peer-review process, where your participation will be assessed by the other members. Information will be provided in class.

Example 1: use data from the thesis of one of the group members or ecological data online (e.g. on dryad: https://datadryad.org) and show how different approaches affect the results (e.g. using transformation vs GLMs, and exploring methods to account for overdispersion).

Example 2: you could find an ecological paper that has an associated online dataset that you can use to try to reproduce the analyses used in the paper and show how analysing the data in a different way would affect the results (e.g. compare imputation to doing list-wise deletion of missing data).

The project will be divided into 3 parts.

i. Outline (5% of class grade)
Submit one outline per group via Canvas. See the schedule above and Canvas for the due date.

One page summary of your project where you present the goal of your final project and the analyses you will perform. Make sure to describe:

- the data, and if the data is taken online, make sure to provide a full citation of the data set and paper
- main questions to be answered
- analyses to be performed

ii. Presentation (16% of class grade: 15% prof's assessment + 1% peer's assessment)
Group presentation in class. See the schedule above and Canvas for the due date.

15 minute presentation (12 min talk, 3 min for questions), where you will discuss:

- goal of the project
- dataset
- analyses performed
- comparisons of the analyses in terms of impact on results and conclusions
- recommendation

While it is a group presentation, you will be graded individually based on the complete presentation and your own presentation skills. Make sure all group members participate and that you help each other create a good overall presentation. Students from other groups will assess your final presentation. More information is available on Canvas.

iii. Written report (20% of class grade)
Submit one report per group via Canvas. See the schedule above and Canvas for the due date.

Similar to the presentation, the main goals of this 10-15 pages (double spaced, including figures) written report are to explain the goal of the project and discuss the pros and cons of the methods explored. You should conclude with a recommendation with regards to the best analysis and the interpretation of the results. The format should include: Introduction, Methods, Results, and Discussion.

- Introduction should explain the main goals of the project.
- Methods should explain the methods explored.
- Results should present the results from all the methods explored and focus on the differences between the methods.
- Discussion should focus on discussing the pros and cons of the different methods, and should make a recommendation on which of the methods explored is the most

appropriate for the data and question. Make sure to give a clear interpretation of the results.

You will be graded on your understanding of the statistical analyses performed and the quality of the written report. For me to be able to assess whether you understand the analyses, you need to clearly describe all of the methods you used, including those covered in class. You need to emphasize why you are exploring specific analyses. Because the quality of the report is also assessed, pay attention to grammar, typos, and paragraph structure (e.g. include topic sentences). Verify that you are clear and concise and that your figures and tables are easy to read (e.g. make sure the axes are written with large enough font to be readable and that the axis titles are easily interpretable). You should write this report as a short scientific publication. Thus, describe your methods with words and equations, not with R code. Similarly, describe your results with words, tables, and figures, not with R outputs. ***I will expect you to make changes according to the written comments I have made after the presentation.*** *You cannot use genAI tools such as ChatGPT to help you complete this report; see details below.*

4. Group participation:
The final project, which is a core component of the class, is a group assignment and many in-class activities are done in groups. Throughout the term, there will be peer evaluations made for you to receive feedback from your peers and assess your contribution to the group work. In brief, you will be asked to write and sign a group contract (to be submitted via Canvas) and then you will be asked to fill two sets of peer evaluations (to be filled via ipeer, see Canvas).

i. Formative peer evaluations
The first set of peer evaluations will not be anonymous and serve as a way for your team to give constructive feedback to other team members and help everyone improve.

ii. Summative peer evaluations (4% of class grade)
The second set of peer evaluations will be used to calculate your group participation grade. You must complete your assessment of other students on time; 10% of your peer grade is deducted per day late. The peer evaluations of your groupmates are due at the same time as the final report.

Missed classes, late assignments, and grade changes:
I do not provide extension for discussion documents and tutorials, as these are associated with in-class activities. If you do not hand-in your assignment on time or miss a class due to valid reasons (e.g., you are sick), the grade will be weighted into your final project (i.e., you won't lose marks, but your final project will count for more). If you have a health-related reason to be late for your final project, I will provide an extension. Otherwise, I will remove 10% for each day past the deadline. Many of the activities are group-based, and thus please be respectful of others and participate equally.

Further reading:

The first set of lectures are based on the book *Ecological Statistics: contemporary theory and application* edited by G.A. Fox, S. Negrete-yankelevich, V.J. Sosa *(available online: http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199672547.001.0001/acprof-9780199672547)*

You are not required to read the associated chapters, but if you find the material challenging, I recommend that you do.

Statement about UBC's values and policies: UBC provides resources to support student learning and to maintain healthy lifestyles but recognizes that sometimes crises arise and so there are additional resources to access including those for survivors of sexual violence. UBC values respect for the person and ideas of all members of the academic community. Harassment and discrimination are not tolerated nor is suppression of academic freedom. UBC provides appropriate accommodation for students with disabilities and for religious, spiritual and cultural observances. UBC values academic honesty and students are expected to acknowledge the ideas generated by others and to uphold the highest academic standards in all of their actions. Details of the policies and how to access support are available here.

Preferred/chosen name: At UBC, students can update their given name to reflect their preferred name across many university systems. Learn more about this process here: Preferred/chosen Names at UBC. If you have any concerns, please feel welcome to reach out.

Use of generative artificial intelligence tools: Over the last years, tools that use artificial intelligence algorithms to generate written content by comparison to a large dataset of training data, such as ChatGPT, have become widely accessible (these are often referred to as "GenAI tools"). Whilst these tools offer many possibilities to support learning, they also come with challenges and are no substitute for one of the fundamental goals of your graduate study, which is to learn through practicing the craft of scientific writing. ***In this class, it is unacceptable to have a generative AI tool write assignments (or parts of assignments).*** See the guidelines from G+PS and the general UBC guidelines, and please refer to UBC's Privacy Impact Assessment for generative AI tools. There are other common applications that use generative AI, including spelling and grammar editors and translation applications (e.g., Grammarly and Google Translate). You may use spelling and grammar editors to double-check an assignment draft or thesis/dissertation. Similarly, you may use a translation application to translate words and short phrases that you have written in another language and wish to use in your assignment. However, it is unacceptable to write your assignment in another language and use an application to translate it to English. *This statement is built on guidelines from UBC, G+PS, Faculty of Science, the Institute for the Oceans & Fisheries, and UBC Science 113.*

We acknowledge that the UBC Vancouver campus is situated within the traditional, ancestral and unceded territory of the xʷməθkʷəy̓əm (Musqueam).