**Statistical Modeling for Ecological Epidemiology**

**Georgetown University**

**Spring 2023**

**1 credit**

**September 1st – December 10th**

**Instructor:** Oswaldo Villena, [Oswaldo.Villena@georgetown.edu](mailto:Oswaldo.Villena@georgetown.edu)

**Office Hours:** TBD

# Course Description

This course will introduce the use of frequentist and Bayesian statistics to analyze epidemiological data (e.g., prevalence data, vector occurrence and abundance, thermal limits for vectors and pathogens). Some of the methods we will study are classification and regression trees, generalized linear models, generalized linear mixed models, generalized additive models. Also, we will fit a diverse set of functions (e.g., Briere, quadratic, polynomial) using a Bayesian approach to estimate thermal performance curves for non-linear data. This course is designed particularly for graduate and upper-level undergraduate students who have had some introduction to R-programming, and it will provide an opportunity to analyze actual population data to learn how to model ecological epidemiologic data. The course will consist of lectures and hands-on practice in computer labs, homework assignments and a final project. R, a free software for statistical computing and graphics, will be used.

# Course Prerequisites

There are no formal requirements, but Introductory Statistics and Introduction to Programming, Data Analysis with R (BIOL 442), or Introduction to Computer Programming for Biologist (BIOL 431) may be beneficial

# Course Requirements

Course assessment will consist of four short take-home programming assignments (15% each), and a final project (40%). For the final project, students could work individually or in groups of 2-3 students. Students will collect open-source data on a topic of interest (e.g., thermal performance curves, occurrence, abundance), analyzed, and reported following a scientific journal template. Students will be graded on the functionality, reproducibility, and documentation of their code, visualization of their output, from analysis, and the reporting of their results. Final project is due by the last day of classes.

# Schedule of Lecture Topics

|  |  |  |
| --- | --- | --- |
| **Week** | **Topic** | **Readings** |
| Week 1 | Course introduction, R software review: Measures of central tendency, measures of dispersion. | 1,2 |
| Week 2 | R software review: data visualization using plot and ggplot functions | 1,2,3,4 |
| Week 3 | Thermal biology of vector-borne diseases: ectothermic insects’ thermal responses, life history traits responses to temperature | 5,6 |
| Week 4 | The Bayesian approach in practice: Fitting non-linear functions | 6 |
| Week 5 | The Bayesian approach in practice: Building thermal performance curves for vector-borne diseases | 6,7 |
| Week 6 | Spatio-temporal epidemiology: estimation of vector’s occurrence using detection/non-detection data; CART models | 8 |
| Week 7 | Spatio-temporal epidemiology: estimation of vector’s occurrence using detection/non-detection data; GLM models | 9, 10 |
| Week 8 | Abundance of vectors: estimation from zero inflated data; use of use of GLMM models (Poisson, Negative binomial) | 10, 11 |
| Week 9 | Abundance of vectors: estimation from zero inflated data; use of GLMM models (zero-inflated and hurdle models) | 10, 11 |

# Course Materials

There is no required textbook for this course. Some recommended resources are listed below

1. R for Data Science – G. Grolemund and H. Wickham (2017) <https://r4ds.had.co.nz/>
2. Advanced R – H. Wickham (2019) <https://adv-r.hadley.nz/>
3. R Data Visualization Cookbook – A. Gohil (2015) <https://cicerocq.files.wordpress.com/2020/03/gohil-a.-r-data-visualization-cookbook-.pdf>
4. Ggplot2: Elegant Graphics for Data Analysis – H. Wickham (2016) <https://ggplot2-book.org/index.html>
5. Thermal biology of mosquito borne-borne disease - Mordecai et al. (2019) <https://onlinelibrary.wiley.com/doi/full/10.1111/ele.13335>
6. Understanding uncertainty in temperature effects on vector-borne disease: a Bayesian approach - Johnson et al. (2015) <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.1890/13-1964.1>
7. Temperature impacts the environmental suitability for malaria transmission by *Anopheles gambiae* and *Anopheles stephensi* - Villena et al. (2022)

<https://esajournals.onlinelibrary.wiley.com/doi/full/10.1002/ecy.3685>

1. Classification and regression trees (CART) - Berk, R. (2008) <https://link.springer.com/chapter/10.1007/978-0-387-77501-2_3>
2. Habitat quality determines patch occupancy of two specialist Lepidoptera species in well-coneccted grasslands - Munsch, T. et al. (2019) <https://www.researchgate.net/publication/329044069_Habitat_quality_determines_patch_occupancy_of_two_specialist_Lepidoptera_species_in_well-connected_grasslands>
3. Beginner’s guide to GLM and GLMM with R - Zuur, A. (2013) <http://highstat.com/index.php/beginner-s-guide-to-glm-and-glmm>
4. Beginner’s guide to Zero-Inflated Models with R - Zuur A. (2016) <http://highstat.com/index.php/beginner-s-guide-to-zero-inflated-models>

# Useful On-line Resources

# R Project. <https://www.r-project.org/>

# R Studio. <https://www.rstudio.com/>

# ggplot2. <https://ggplot2.tidyverse.org/>

# R-bloggers. <http://www.r-bloggers.com/>

# R Graphical Manual. <https://www.imsbio.co.jp/RGM/R_image_list?page=1260&init=true>

# R for SAS, SPSS, and Stata Users. <http://r4stats.com/>

# Attendance Policies

Illnesses and absence: Note that this is a small course with few scheduled meetings. Absences will be problematic because the classes will build on one another. Lecture notes will be posted on CANVAS. There may be times when you are sick or have to travel; please contact the instructor to make arrangements.

# Student Accommodations

Student Accommodations: Students should speak with their instructors before or during the first week of classes regarding any special needs. Students can also visit or contact the Academic Resource Center ([arc@georgetown.edu](mailto:arc@georgetown.edu)). Students seeking academic accommodations should register with Services for Students with Disabilities (SSD). SSD arranges reasonable and appropriate academic accommodations for students with disabilities. Please visit <https://academicsupport.georgetown.edu/disability/> for more information on student accommodations.

# Letter Grades

Final grades are rounded up to the nearest integer, then assigned as follows:

|  |  |
| --- | --- |
| A | 95-100 |
| A- | 90-94 |
| B+ | 87-89 |
| B | 83-86 |
| B- | 80-82 |
| C+ | 77-79 |
| C | 73-76 |
| C- | 70-72 |
| D | 65-69 |
| F | <65 |