Exam 3 – Cumulative Final Exam

ENTO 8900: Advanced Quantitative Analysis

Due December 8 – 11:00 AM (Submit to eLC)

You will be given access to various datasets and the requisite background information to answer questions and perform the analysis to the described specifications. Please read the instructions carefully, and do not hesitate to contact me for clarifications. You are permitted to collaborate with other students to complete the exam, but you are encouraged to make a good-faithed effort on your own at first. This will make sure that you maximize your learning and retention of the material. However, you should endeavor to submit answers in your own language as concerns related to academic honesty will apply. You should submit answers to exam questions as either a word document with imbedded R script/console code/output or as a R markdown file (your choice). Files should be uploaded to eLC by the listed deadline to earn full credit (see syllabus for late policy). Make sure to provide details related to code for uploading, managing, and transforming data.

Access datasets for this exam via the course Github folder: (“https://raw.github.com/Conorfair/ENTO\_8900/blob/main”) and open the raw data to select the appropriate URL to use in your RStudio script code.

1. Describe two out of the five Gauss Markov assumptions. Explain why time-series data fail to meet one or more of these assumptions. **3 Points**
2. Describe (in detail) the two motivations discussed in class for the use of random/mixed effects modeling. **2 Points**
3. The follow data are to be analyzed using an event history model. Follow the modeling process explained in class and assess the impact temperature and habitat quality has on the time it takes for an adult stonefly to emerge from a stream. Researchers believe that as temperature increases the rate of emergence increases, and that this relationship might be different for streams of good and poor habitat quality. Produce both a Kaplan-Meier and predicted survival function from the Cox Proportional Hazard model. Interpret the results from the analysis in the broader biological context. **10 Points**
4. Carefully consider the dataset for this question. Given the type of community data, what kind of measurement of association (Sørensen Coefficient, or Bray-Curtis) would you use for an NMDS analysis? Complete the NMDS analysis with your chosen measurement of association. Produce a 2-D ordination plot with the relevant experimental design information, and annotations with the necessary statistical reports. Finally, perform an ANOSIM and interpret the results within the broader biological context. **15 Points**
5. Your colleague has reached out to discuss what analysis would be best given their questions, hypotheses, and data collected. The details are given below:

Data were collected from twenty different urban gardens surrounding the greater Athens area. Aerial netting collections were completed equally among the twenty different gardens twice each month (at least 10 days apart) from May through September. Data are pooled over the growing season (ignore possible auto-correlation among error structure). Currently, the number of bees collected during each collection event are recorded. There are plans to identify each bee to species to broaden the analysis. Information regarding abundance, species richness, diversity, and community composition are of interest to describe the differences between the twenty different gardens. Speaking of the gardens, there are two variables of interest we used to describe the different gardens. 1) is the species richness of the different floral resources planted in close proximity to the cash crops grown on site (range is from 0 to 10 different species). 2) is the presence or absence of nesting habitat for pollinators on site (bee hotels and bare/undisturbed soil). Five gardens have included bee hotels, five gardens have established locations where soil will remain undisturbed and act as nesting habitat for soil nesting bees, five gardens have adopted both approaches of nesting habitat, and five gardens have adopted neither as a control. Our goal is to test how pollinators respond to increases in floral resources and how this relationship interacts with different types of nesting habitat types/availability. We hope to understand what resources (floral/food or nesting) are limiting to pollinator communities found in these urban gardens.

Walk them through the process/ethos dictated in Beckerman et al. 2017.

1) Clear question and right data – discuss the types of data included in the dataset and describe how these types of data would influence the analysis and figures

2) Inspect/visualize data to answer your question – carefully consider how the data were collected; what kinds of follow up questions would you ask; what initial data visualizations/figures would you want to review?

3) Statistical model that reflects your question – given the above information, provide a suggested analysis

4) Check the assumptions of that model – for each analysis you suggest, provide a list of the assumptions that should be met and how you would test them

5) Interpret the model; confirm/refute your hypotheses – explain in abstract how you would interpret possible outcomes given the model(s)/analyses suggested above

6) Clearly communicate your answer in a figure – sketch or describe potential figures that you would want to build to help communicate your findings

**20 Points**