Preliminary exam questions

*for*

Martin Simonson

16 March 2020

Prepared by Stephen J. Dinsmore

**Instructions**: You have the entire week (8 a.m. Monday to 5 p.m. Friday) to complete these four questions. As we discussed, you are free to use any resources that may help you answer the questions (e.g., open book), although you are not to talk to anyone else during the exam. I’ve tried to provide some general guidance on the length of your response to each question, although you are free to write as much as you feel is necessary. Good luck!

1. One general perception in our profession is the notion that research without management implications (e.g., theoretical ecology) is of diminished value. I would like you to contrast theoretical and applied ecology, citing specific examples of each from your own areas of research. Should our profession continue to emphasize applied ecological questions, or is some other approach more worthwhile? Your response should demonstrate the breadth of your knowledge in these areas. [~5 pages]
2. Most sampling strategies are comprised of two primary elements: an estimator and a sampling design. Together, these two elements can be used to attain a desirable balance between precision and bias. Focusing on bias, the three primary sources are statistical bias, selection bias, and measurement bias. Provide a definition and example(s) of each and then rank and justify these biases according to *your view* of the extent to which they have, or currently are, detracting from the quality of data guiding fisheries science. [2-3 pages]
3. Below is a 2x2 contingency table showing the observed breeding status (B = breeding, NB = not breeding) for a hypothetical fish in a given year. Furthermore, there are two classes of fish in each breeding category: fish attempting to breed for the first time (First), and experienced breeders (Exp).

B NB

First 110 45 155 (%B = 71%)

Exp 92 63 155 (%B = 59%)

Total 202 108 310

* 1. Create a partition of the above table into two possible 2x2 tables (which must sum to the numbers in the above table) that constitute a good example of Simpson’s Paradox (note: there is no unique solution to this).
  2. State Simpson’s Paradox in its general form. Why does it raise concern regarding the analysis of observational ecological data, even under a model-based data analysis?
  3. Why are we nearly immune to Simpson’s Paradox when using a well-designed and highly replicated experiment?

1. Please complete a short analysis of the seasonal survival of a fish (species is irrelevant) in a hypothetical lake system. I have attached two files (.dbf and .ftp) that are formatted for Program MARK. I am interested in seeing your approach to this type of analysis, with less emphasis on a “right” answer. Also realize that this is a “real” analysis in the sense that the data might not be perfect, all parameters might not be estimable, etc. Please begin by providing a *brief* justification for the factors that you considered in your analyses, and your general approach to building models for consideration. Then run an appropriate set of models to answer the questions below. I’ll want to see both of your MARK files, so please include them with your response to this question. Please also label the columns in the design matrix and model names in a way that I can understand them.

Here is a little background about the (fake) dataset. First, carefully read the documentation at the top of the input file. Note that there are 20 occasions (corresponding to April and September sampling periods in each of 10 years) and four individual covariates in the input file. The fish were tagged and recaptured during two 1-month-long (April and September) periods using standard sampling techniques. There are no missing data. Assume that all recaptures were measured without error. Here is a little more information about the covariates:

* **Sex** codes for whether the fish was a male (1), of unknown sex (0), or a female (-1). There is zero chance that they switch sexes during the study.
* **Length** is measured at capture as the total length from snout to tip of folded caudal fin in cm.
* **Mass** is measured at capture in kilograms.
* **Scaled white color** is measured from a digital photo of each fish, and is essentially a proportion of the skin (ranging from 0 to 1) that was white. It has been suggested that paler fish may be more susceptible to harvest.
* In addition, the annual harvest of this fish (in metric tons) is 300, 330, 400, 440, 600, 525, 540, 510, 575, and 560 for the 10 years. Harvest is in fall.

When you open the .dbf file in MARK, you will get the message “No output is stored in this Results Database”; click o.k. and you will be directed to an empty results browser where you can begin building models. You are free to email me if any of this is unclear. **And to keep the analysis on track, do not consider (or even run) the fully time-dependent model.**

After you complete your analysis, please answer the following questions:

* 1. What is your best estimate of the seasonal survival of fish (April to September, and September to April)? Be sure to show your work.
  2. Is there any evidence that sex, length, mass, or scaled white color is an important predictor of survival or capture probability in this population? If so, then how and under what conditions? (1-2 pages)
  3. What problems, if any, did you encounter in your analysis and how might they be resolved? (1 page)
  4. Summarize all of the findings from your analysis in a short (<400 words) paragraph, just as you would write for an abstract of a manuscript.