Stat 534 Spring 2023 - takehome exam

Due: Wednesday, 13 December, to canvas by 11:59 pm.

There are three questions, each with multiple parts.

FYI: I try my best to grade anonymously. Please put your name only on the signature page. Do not put it on any other page.

You are allowed to use your book(s), your notes and pre-existing online resources. **You are not allowed to ask a question to classmates, friends, or an online forum**. Direct your questions to me and no one else. I will not provide code, but I am very happy to answer questions about syntax, output, or errors. If you do not understand a question or what it is looking for, ask for clarification.

Please do not include the questions with your answers. Just write the question number, e.g. 1c) then your answer.

Many questions ask you to describe the analysis used to answer the question. This can be using words or words and the specific R code used for that question. Please include the entire workflow for a question after the signature page. The workflow should include the code to read in the data, do any data manipulation steps, and do all the analyses,

Arrange your answers in the following order:

Answers to question 1, answers to question 2, answers to question 3,
signature page (or copy of it if you're using Word)
R workflow for Q 1, R workflow for Q 2, R workflow for Q 3

Reminders:

1. **You are not allowed to ask classmates or friends for help. Big difference from HW assignments!** Please ask me, not a classmate, for help if you don't understand the question or are having computing issues.

2. Please sign the appropriate line on the honesty page when you are finished working on the exam.

3. Don't spend more than 30 minutes trying to fix a computing issue. If you haven't fixed the problem in 30 minutes, e-mail **me** for help.

Name: _____

Reminder: You are allowed, and do not need to report, use of pre-existing online resources, so long as you don't post a question to one.

Please check the appropriate box, sign, and include at the end of your answers.

☐ I completed this exam without assistance from friends or classmates.

☐ I received the following assistance from friends, classmates, or posts to online forums (please describe below).

Signed: _____

1. This question examines data from a fish survey on the Missouri river. Fish were sampled at two habitats (chute, main) in each of 3 locations (C, L, and O) along the river. Main habitat samples are from the main stem of the river; chute habitat samples are from side channels with shallower water. There are between 1 and 3 samples at each combination of habitat and location. These data are a subset of a much larger data set that included data from four other sampling methods; you will only consider the data from the large hoop nets (gear = LH).

   The first three questions concern the subset of data from two locations, L and O, and one species, freshwater drum, "fwdm". These are counts of the number of that species caught in that sample. Treat these data as if each row was a simple random sample from Poisson distribution with a location-specific mean. These data are in fwdm.csv.

   (a) 5 pts. Use maximum likelihood to estimate the mean counts for the L and the O locations. (Your answer should have 2 estimates)

   (b) 5 pts. Test the null hypothesis that the two locations have equal mean counts. Name the test statistic you are using (T, F, Z, Chi-square, something else), report the value of the test statistic, and report the p-value.

   (c) 5 pts. Estimate the ratio of mean counts and a 95% confidence interval for that ratio. These will allow you to complete the sentence:

   The mean count of freshwater drum at L is an estimated _____ times that at O,

   with a 95% confidence interval of (_____, _____).

   The following questions concern all species in the 12 samples from 3 locations and both habitats. There are 8 species found in one or more samples. You most interested in the difference in species composition between main channel and chute habitats. Your analysis will start by evaluating the interaction between location and habitat. If that's not significant, you will drop the interaction and evaluate the difference between habitats, averaged over locations. If the interaction is significant, analyze each location separately. Use functions in the mvabund library for your analysis. Fish counts are in fishLH.csv; sample characteristics are in fishenvLH.csv.

   (d) Is the difference between chute and main habitats similar at the three sites? Or, is there evidence of an interaction between location and habitat?

      i. 5 pts. Briefly describe the analysis you used to answer this question

      ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

   (e) Is the species composition different in chute and main habitats?

      i. 5 pts. Briefly describe the analysis you used to answer this question

ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

2. The data in critter.txt are from a mark-recapture study of a closed population with 10 trapping occasions. You are primarily interested in estimating the population size. You aren't sure what might be an appropriate model for the detection process. You only have capture history information; you do not have any other information about each individual.

   (a) Consider the M0, Mt, Mb, and Mtb models. For Mtb, use a model that is not overparameterized. What is the most appropriate model for the detection process?

      i. 5 pts. Briefly describe the analysis you used to answer this question

      ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

   (b) You suspect that individuals vary in their detectability. In other words, at any capture occasion, some individuals are more likely to be detected while others are less likely. This is true both for the first time an individuals is captured and for recaptures. Consider variations on the M0, Mt, Mb, and Mtb models that incorporate individual variability in detectability. What is the most appropriate of these models for the detection process?

      i. 5 pts. Briefly describe the analysis you used to answer this question

      ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

   (c) Evaluate whether it is necessary to account for heterogeneity in detection probability.

      i. 5 pts. Briefly describe the analysis you used to answer this question

      ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

   (d) 5 pts. Using the most appropriate detection model, estimate and report the population size.

   (e) 5 pts. Estimate a 95% confidence interval for the population size.

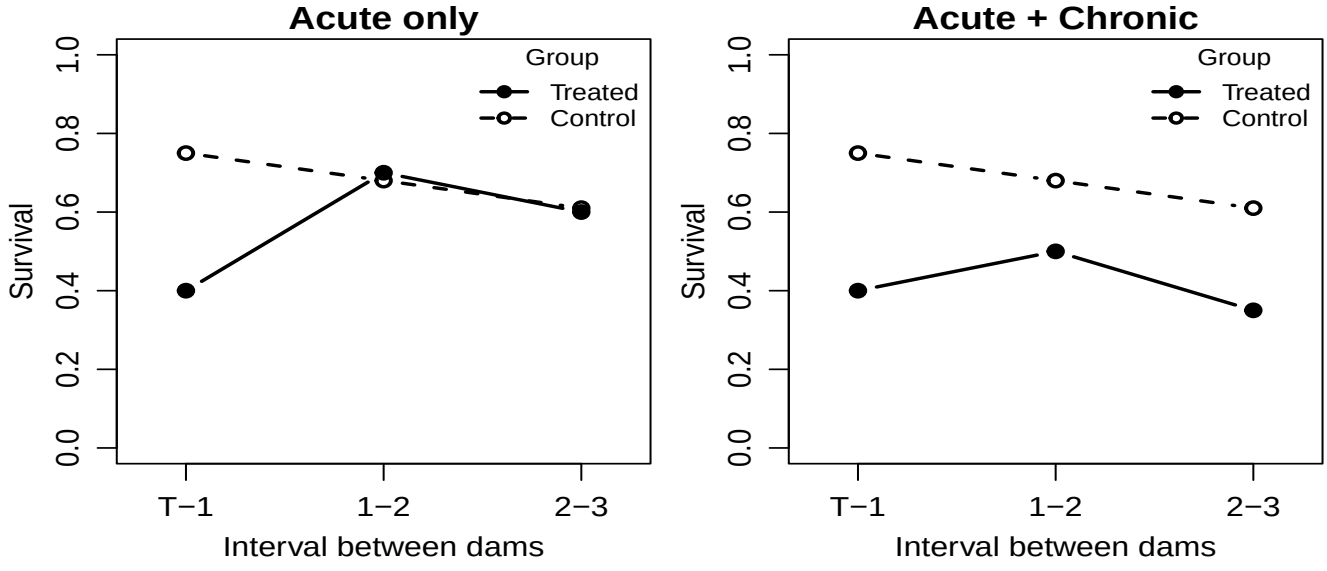
3. The data in salmon.txt were collected to evaluate mortality of salmon passing through a turbine in a hydroelectic dam. The background for this study is a river with hydroelectic dams. Historically, salmon spawned in the headwaters and juvenile salmon (smolts) traveled down river to the ocean. Now, water flowing through turbines at the dams turns generators that produce electricity. Salmon also have pass through the turbines. These data were collected to estimate mortality due to passage through turbines.

   2000 smolts were individually tagged with PIT tags and randomly assigned to treatment or control groups. The 1000 treatment (trt = 1) smolts were released immediately upstream of the dam and had to pass through its turbines. The 1000 control (trt = 0) smolts were released immediately downstream of the dam and did not pass through the turbines. Individual fish were detected as they traveled past each of 4 downstream dams. Time 1 is the release location, time 2 is the first downstream dam, time 3 is the 2nd, etc. Note that detection was far from perfect, for various reasons, and some fish were never detected after release.

   Some assumptions are given as starting points for analyses. The first three are based on the setup of the detectors at each downstream dam. It is possible that:

   - 1) Each of the four downstream dams may have a different detection probability.
   - 2) Detection may be modeled by a logistic linear function of downstream dam number, i.e., time.
   - 3) Treatment and control smolts may have the same detection probability at each dam.
   - 4) Survival may differ between each downstream dam.
   - 5) Passage through a turbine has both acute and chronic effects. The acute effect is the mortality due to passing through the turbines at the "treatment" dam. A chronic effect is the increased mortality of treatment smolts further downstream.

   These plots illustrate the difference between acute and chronic effects of turbine passage. The treatment dam is labeled T; downstream dams are numbered sequentially.



The analysis has two goals:

- To estimate the effect of the turbine on smolt survival
- If there is an effect of survival, is there a difference only at the first dam (an acute effect) or does turbine passage reduce survival further downstream (a chronic effect)? Note that once past the first (treatment) dam, both groups of smolts encounter the same conditions.

(a) Use the data to assess assumptions 1), 2) and 3).

   i. 5 pts. Briefly describe the analyses used to answer this question

   ii. 10 pts. What are your conclusions? Support your conclusions with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

(b) Is there an effect of the turbine on smolt survival between the treatment dam and the first downstream dam? If so, how big is that effect?

   i. 5 pts. Briefly describe the analysis you used to answer this question

   ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.

(c) Is there an effect of the turbine on smolt survival after the first downstream dam? If so, how big are those effects.

   i. 5 pts. Briefly describe the analysis you used to answer this question

   ii. 5 pts. What is your conclusion? Support that conclusion with appropriate numerical results. These could be estimates, confidence intervals, p-value(s), or a graph depending on what you feel most appropriate.