Exam for Statistics for Biologists/BIOS14, ht 2021

Instructions

Do this exam when you find the time and hand it in no later than **2022-01-16**! For each section (or question within a section) I need some text that you write in answer to the question, as well as some output from R so that I can see how and why you obtained your answer. A few sentences or a short paragraph per question is usually enough. The exam is composed of 4 sections worth 10 points each. A minimum score of 24 points (60%) is necessary to pass.

To complete the exam, you may obviously use any of the materials provided during the course, or any literature you may need. The only requirement is that you solve it on your own. If I have formulated the questions unclearly, or you have other questions, contact me for clarification. You are free to ask pretty much any question, but I might answer in general terms only.

Note that the text you write is at least as important as the output! Sometimes there is more than one way the data could be analysed, but that one method is better than the alternative(s). If you do an analysis that isn't the best possible one, but which isn't entirely incorrect, then I might be able to give you partial credit if you explain clearly why you chose that particular analysis.

If you want to use R Markdown in Rstudio to prepare your answers, that's fine. Otherwise, you can copy the text from the R console window (which will include both your input and R's output) into a word file or other text editor. Then add your comments and answers to the questions to the same file. Just make sure to indicate what parts are input/output and what parts are your comments. Don't forget to include any figures in the file as well.

Create **one answer file per section in pdf format** and submit them via Canvas. You can also include files with your R scripts, but this is optional.

Some general advice (not only for this exam)

- Always plot your data! A visual inspection is a great help in determining what kind of model/test/transform might be appropriate for the problem at hand. Also, plot in several different ways to "get to know" your data properly.
- Don't forget to test the underlying assumptions of whatever test you use. **This goes for all questions on the exam, whether it is asked for or not!**
- In an ANOVA-style model, don't forget to decide which factors are fixed and which are random!
- Beware of pseudoreplication!
- Contemplate **beforehand** what statistics to use, to assure your experiment and/or measurements are properly designed and sufficient to answer your questions.

Good luck! ☺ Jessica

Question 1

Dr. Webb is an aquatic ecologist who is interested in trophic interactions between species. She works on lake and wetland ecosystems in Scandinavia, and has recently carried out studies on a number of different types of species. We will help her analyse her data.

Use the data file plankton.csv. It is data on the density of a plankton species sampled in 40 different small lakes. These lakes were part of a grand scale experiment. Half of the lakes had been experimentally treated by removal of some fish species. The idea behind the experiment is that it may be possible to influence how the density of plankton increases with available phosphorus concentration, depending on whether fish have been removed or not.

a) Test if the (linear) relationship between phosphorus concentration and plankton density depends on the treatment. (3p)

b) Illustrate your data with an appropriate graph. Motivate your decision. (2p)

c) Write out the linear equations for the model you analyzed in a. Comment on what conclusions you can make from the experiment. Specifically, how do you interpret the model parameters and their confidence intervals? (3p)

d) Would your conclusion about the effect of fish removal have been different if you did not take the effect of phosphorus on plankton density into account? Why or why not? (2p)

Question 2

Eutrophication of lakes and wetlands due to agricultural runoff is a problem for ecosystem functioning, so Dr. Webb has a collaboration with a forestry researcher to see if it's possible to stop some of the agricultural fertilizer ending up in her wetland study sites. They carried out an experiment to see if runoff water from fields could be captured and successfully re-used to fertilize pine trees.

Use the data file pine.csv. This is data from an experiment in which pine seedlings were grown for one year. The experiment was carried out at 10 randomly chosen study sites (`site`). Seedlings were planted and exposed to two different fertilizer treatments (`treat`) – one is a commercial fertilizer solution, and the other is captured agricultural runoff water. One group of plants was left as a control, without any fertilizer added. The lengths of the seedlings were measured after the year had passed. Each row in the dataset represents a mean of 5 seedlings.

a) Test if there is any effects of the treatment. Does captured agricultural runoff water seem to be an effective fertilizer for pine trees? (3p)

b) Make a graph that illustrates the analysis you have carried out in part a. Motivate your choice of graph type. (2p)

c) Discuss the results of tests of the assumptions that your analysis is based upon. (2p)

d) State the null-hypotheses for a and c. (3p)

Question 3

Because she is interested in trophic interactions, Dr. Webb also wanted to know how the presence of various host plants influenced the presence of some insect species at her study sites. Use the data

file moth.csv. This is data on the occurrence of two sub-species of a moth, collected at 30 different sites. At each site the sub-species found is determined and the habitat is classified as being warm or cold (`hab`). However, it is suspected that one of the subspecies (`subs=1`) prefers a particular plant species for egg laying, whereas the other sub-species is known to be a generalist (`subs=0`). The density of this plant is also given (`host`).

a) Test if the occurrence of the two subspecies differs between habitats (ignoring the host plants). Motivate your choice of test. (3p)

b) Test if the occurrence of the sub-species is determined by the availability of the potential host plant, as well as by habitat. How do you interpret your results? (4p)

c) Illustrate your results from parts a and b. (3p)

Question 4

Finally, Dr. Webb was curious about the predator species around her study sites. Use the dataset morph.csv. This is morphological data from different female morphs of a species of damselfly from a single study site in a single year. We are interested in whether the female morphs differ in size, since this might influence what prey they prefer to capture.

a) The five morphological traits are highly correlated with each other, so rather than test each trait separately, we would like to obtain a new variable that somehow combines the information from all five traits into a single measure of overall size. Calculate such a variable and explain why you used that particular method to do so. (4p)

b) Test if the female morphs differ in overall size. Does it seem likely that they eat different types of prey? (3p)

c) How could the design of this study be improved? (3p)