

BIOL4062 & 5062: Analysis of Biological Data

Instructor: Henry Whitehead
E-mail : HWHITEHE@DAL.CA

Biologists are increasingly using quantitative techniques to analyze larger and larger data sets. It is clear that a command of the available analytical techniques is an important part of the set of skills which are expected of a trained biologist, especially those working in the broad area of ecology. The class will introduce techniques available for the analysis of biological data, including correlation, regression and multivariate methods. Emphasis will be on the practical use and abuse of these techniques rather than derivations or mathematical formulae. Students will explore real and realistic data sets. There will be lectures on the use of techniques, and students will try them out using real and simulated data sets.

Type 1 assignments

In these assignments, each student will be given an artificial, but realistic, data set to analyze. The data sets given to the different students will be structurally similar but numerically different. Graduate students (BIOL5062) will sometimes be asked to perform more analyses than undergraduates (BIOL4062). The data should be analyzed using a computer statistical package (see below), and you should hand in a short write-up of each (more detailed instructions with each data set). In the write-up explain clearly what you did, what you found, and what you think the results might mean biologically.

Type 2 assignments

Each student will find a biological data set, and then analyze it. The analysis:

Should not be part of your past, present, or future Honours, MSc or PhD thesis, or used for another class—this is self-plagiarism. However, you can use the same data set, or part of the same data set, as in a thesis or another course, but do a totally different analysis on it.

Should not be that, or repeat that, done by someone else—this is plagiarism.

Can use a data set collected by your supervisor, or someone else, but you should ask them about using it.

Can use a data set that you find on the web, or somewhere else, but you should check that it is OK to use it.

Can later be submitted for publication, but you must check that you have all necessary permission from those who collected the data.

Must address at least 3 biological questions (undergraduates; BIOL4062), or 4 questions (graduates; BIOL5062)

Must be of a data set of at least of the following sizes (ask Hal for exceptions or in case of uncertainty):

>50 units x >3 variables, for undergraduates (BIOL4062)

>50 units x >5 variables with at least 2 types of variables .(e.g. “Dependent” and “Independent”; “Species and “Environment”) or link two data sets with one at least as large as the undergraduate data set, for graduates (BIOL5062)

The analysis of this data set is done in 4 steps (2a, 2b, 2c, 2d).

2a. You will have a short meeting to discuss your proposed data set and proposed analysis (bring a draft of 2b assignment). Hopefully you will get useful feedback.

2b. Description of data set and proposed analysis. Hand to Hal a brief description of the data set, with the following information: where it came from; its structure(s) (number of variables, units, names of variables, types of variables, ...); proposed biological questions; proposed analytical methods; possible problems. The mark that we give will depend on you presenting this information and on whether the data set seems suitable to look at interesting biological questions, and not on whether, at this stage, you choose exactly the best analytical techniques—we will help you with this. The purpose of this step is to make sure you are setting off on a suitable track.

2c. Presentation of your results to the class by graduate students. 2c (i). Graduate students present the biological questions being addressed, a brief description of the data set, how you analyzed it, and your conclusions. Hopefully you will get useful feedback. We will give an example of this kind of presentation later in the class, or see attached. Undergraduate students should go to these presentations and will be tested on general issues arising from them on the last day of class (2c (ii)).

2d. Write-up of your analysis as for a scientific journal paper—no more than 5 pages (BIOL4062) or 7 pages (BIOL5062) of single-spaced text, not including references, tables, figures. Explain the biological question, your methods in sufficient detail for someone to replicate them, any problems you had, and your biological conclusions. Show graphically, or in tables, the major effects you found. Do not just present summaries of ordinations or significance levels of hypotheses tests. The Introduction and Discussion can be shorter and less detailed than you would ordinarily write in a paper. However, they should be sufficient to give a good feel for the biological issue being analyzed and the potential biological significance of the results.

General assignment information

Additional information : Written assignments (1a, 1b, 1c, 1d, 1e, 2b, 2d) must be handed in to me, or put in my mailslot in the Biology Office, by 16:30 on the due date. 10% off for each weekday late, without official medical excuse. You may hand in assignments double-sided or printed on one side of previously used paper.

Plagiarism

Do not plagiarize (including self-plagiarize)! I catch plagiarizing students quite often. If you are uncertain of what is and is not acceptable, ask me, or see

<http://www.registrar.dal.ca/calendar/ug/UREG.htm#12>