Subject name

Assignment 1: Designing data collection

Planning your sampling program efficiently means that you must make decisions about the kind of data that you must collect to answer a particular biological research question. Your aim is to collect data that allows you to answer the question unambiguously. Often you will have several competing answers, and the logic of your sampling program should allow you to distinguish between these answers. You also need to collect enough samples for your statistical analysis to be able to separate the alternatives. Collecting enough samples can be difficult. All research programs are ultimately constrained by time and money, particularly when the research is part of a degree. You will be limited by the time available to complete your degree, and the resources available in your research group. In some cases, there may be other constraints, such as the supply of material; there may be limits in numbers of animals bred in animal house facilities, greenhouse space, or volume of tissue samples that are provided. These constraints determine your overall effort, and research projects generally have several components. You will need to decide how to allocate your time and resources between these components. If you direct too little effort into one component, you may not get an answer to the question, but if you put too much effort into one component, you may not be able to complete other components satisfactorily.

Planning your sampling can make use of tools such as power analysis, but to use these tools, you need to have a good idea of what you will do with the data when you have them. Our emphasis on effective planning has a side benefit of having you think about data analysis early in the process, rather than collecting data and then thinking about what you can do with the data.

This assignment requires you to plan a sampling program in the area of your research. It is designed to get you thinking about how you will collect some of your own data, and some of the situations that are particular to your research interests. You’ll find it most helpful if you can directly apply some of the techniques that we’ve talked about to your own project, but that may not be the case at this stage of your degree. You may not be well enough advanced in your research to be collecting data, or what we’ve covered may not match perfectly your particular research. Even so, some topics covered may aid your understanding of the methods you will eventually apply. For example, if your project area was systematics, you may find yourself constructing phylogenies. The phylogenies might be constructed using particular pieces of software or approaches, but sophisticated multivariate analyses underlie the construction of phylogenies. Fundamental to these approaches is measuring the similarity between samples, using a suite of measurements (morphology, molecular data, etc.).

You should complete the sections below of this document, and then upload your assignment through Canvas. You should work on your own to complete the document. We will meet in small groups to discuss the proposed sampling programs.

**The marking scheme shows the importance of each section of your answers;** use this to guide you on how much detail you’ll need for each section. **The marking scheme does not indicate the number of words that you might need to write. Some important questions may only need a few words to show you understand!**

Word limit: 1500 words

Due date: **30 July, midnight**

|  |
| --- |
| What (biological) question do you hope to answer with your data? (1 mark) |
| Translate this question into a statistical query (hypothesis, prediction, etc.) (2 marks) |
| Describe the structure of your sampling (3 marks) *Outline the groups, etc., in your data. For example, do you have discrete groups, corresponding to experimental treatments? Are there other natural groupings in your data? Will your data collection be spread along a gradient of some other variable? Are there other possible sources of variation in the data that you need to eliminate by the way you arrange your samples?* |
| What kinds of data will you collect? (1 mark) *Outline the kinds of measurements that you will make, including whether the variables are continuous, etc.* |
| What are your experimental or observational units (i.e., the replicates)? (3 marks) |
| Describe the randomization procedures that you will use (2 marks) |
| What statistical model will you fit to your data? (3 marks) |
| What are the assumptions of this model? (2 marks) |
| How will you decide the number of samples required? (3 marks) *Provide details here, rather than just a broad description of a technique. You should think about the information that you would need to put into a power analysis or other relevant approach* |