Final Examination Part I

Biology 697: Introduction to Computational Data Analysis

11/27/2012

Part I of the Exam: Below is the description of a research scenario. It is general, so feel free to design whatever you would like. What I'm looking for is a well designed set of studies (between 1-3 experimental or observational studies). The design should be fully justified with respect to the types of models being used, the error distributions chosen, and the relevant sample sizes reflecting a concern for minimizing type II error. Note, while you may not be able to plan a priori for specific effect sizes and variances, you can often approximate relationships assuming that you have z-transformed variables. You can easily back transform them later when you have real data and re-evaluate your assumptions regarding power, etc.

In short, take the below information, design some meaningful studies, and justify them. Make sure that I have enough specific information specified as clearly as possible so that I can plug in numbers into a few equations I have written that describe the system, and hand you back data.

The Problem at Hand:

You're interested in studying the dynamics of lake communities with a tricky species if *Daphnia* that exhibits some funny evolutionary behavior. This little water flea has multiple copies of a gene that ends up coding for spines in its genome (1-20 copies). Having many many spines can be an advantage, as it's less likely to be eaten. But, you think that also may mean it is a less efficient forager. In different lakes, the number of genes in local populations seems to center around some optimum number.

But the *Daphnia* are but one piece of the tangled web of interactions in this system. Your advisor hands you this conceptual diagram of how the dynamics of these lakes work - covering everything from lake size to nutrient concentration to the abundance of frogs and insects. It may or may not be the entire web of interactions, but it's the best you all have currently. In the network, any arrow that connects two boxes represents an effect - could be linear, could be nonlinear. Any arrow that intersects another arrow represents an interaction effect.

She charges you with answering two questions along the following lines
1) How strong are top down (predation) versus bottom-up (nutrient & food supply) forces in this food web. Feel free to chose any compartment.
2) How does genetics modify the outcome of top-down processes?

You have \$2000, a state full of lakes, a set of pond mesocosms, two trained techs, and two new undergrads. Design studies that will address these questions within your budget. Here's how things cost...

Observational Studies: To get to any lake to sample it is \$5 in gas (on average). Given rates of pay per time, it's \$2 per sample of 1 variable at a lake in the field for an undergrad, but the amount of measurement error may be high. It's \$5 per sample of 1 variable for a field tech, but the amount of measurement error will be reduced.

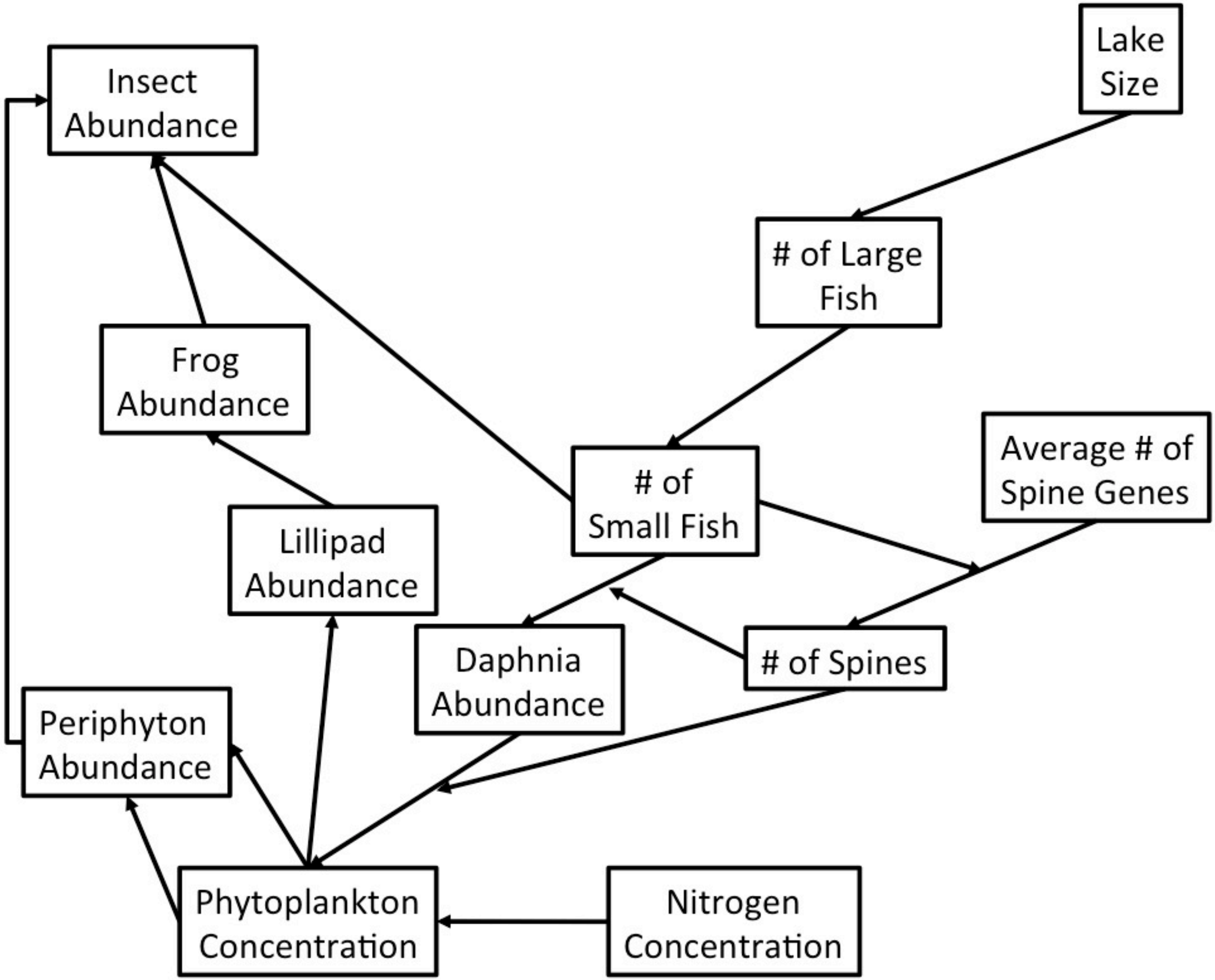
It's \$3 extra for determining spine counts from *Daphnia* samples, assessung nutrient, or phytoplankton concentration due to work back at the lab needed to assess these quantities. Add \$7 extra per sample for any genetic analysis to get # of spine genes per individual in a sample. Knock off \$1 per sample if you use undergrads, but, residual variance will be increased.

Experimental Studies: Experiments are great, but more expensive. For every mesocosm you choose to setup, there's a \$30 startup fee (plumbing, man hours, etc.). Manipulating things that could be sampled cheaply before (abundances) is, again, cheap! \$5/item/replicate. Manipulating something like nutrient concentration or phytoplankton concentration is \$10/item/replicate. And manipulating something fancy, like the genetic structure of a treatment population, is actually quite difficult due to the need to culture strains, etc. \$30/item/replicate.

Some useful information:

- Lake Size: varies from .05 - 15 km^2.
- # of Large Fish varies from 0-0.5 per m^3.
- # of Small Fish varies from 0-15 per m^3.
- # of Spines per Daphnia in a lakes varies from 0-50 and is Poisson distributed in any lake.

- Daphnia abundance ranges from 10-300 per liter.
- Nitrogen Concentration ranges from 0-700 μg per liter.
- Phytoplankton concentrations range from 50-300 mg per liter.
- Periphyton abundance ranges from 0-100% cover in a 25cm^2 area.
- Lillipads are Poisson distributed in lakes and vary from 0-25 per square meter.
- Insects caught on sticky traps around lakes vary from 15-1500 per trap.


Lake Size
# of Large Fish
Average # of Spine Genes
# of Small Fish
# of Spines
Insect Abundance
Frog Abundance
Lillipad Abundance
Daphnia Abundance
Periphyton Abundance
Phytoplankton Concentration
Nitrogen Concentration