Introduction to course

If you are taking this course, I gather that you have (or will have) data in hand, and you are interested in drawing some inferences from them. For example, you might be interested in quantifying the magnitude of a treatment effect, or the rate of population change over time - these are examples of **parameter estimation**. Alternatively, you might want to test whether the treatment effect differs from a control, or whether the rate of population change is different from a hypothesized value - these are examples of **inference**. In statistical modeling, we start with the data, and we ask "*what can we say about the cause of the data - i.e., the process(es) that generated these data?*" Armed with a deterministic model of the process(es) giving rise to the data, we then incorporate stochasticity to account for uncertainty in our model, observations, or both. The workhorse under the hood of any statistical model is probability theory. We choose a reasonable probability distribution for our response variable and estimate the value of unknown parameters using an appropriate engine (e.g., ordinary least squares, maximum likelihood, Markov chain Monte Carlo).

Even though this course is concerned primarily with statistical estimation and inference, we'll need to brush up on some basics of probability theory. In probability theory, we think about processes that generate data, and we ask "*what can we say about the data generated by such a process?*" In other words, we start with the cause (probability distributions, model, parameters), and then we can generate the effect (data). We often talk about data-generating processes in a modeling framework. So, we can define probability theory as the study of data generated by specified processes.

This course is targeted for graduate students in the biological and environmental sciences, but students from any discipline are welcome. It is important to have at least some background in introductory statistics and scientific computing with R - I will assume familiarity with these topics. Although certain calculus concepts are important in statistical modeling, calculus derivations are not. It turns out that numerical (rather than analytical) approaches are necessary for all but the simplest model scenarios.

We will take a Bayesian approach in this course. Why Bayes? There are many compelling reasons, not least of which is that Bayesian methods are becoming standard in the life and environmental sciences. At minimum, students should be able to understand this modern approach to statistics in the literature. Moreover, the Bayesian philosophy offers an intuitive way of speaking about the probability of parameters. No more fussing about with the interpretation of a p-value or limiting yourself to a framework of null hypothesis testing. Pedagogically, learning statistics in a Bayesian framework allows us to peak under the hood, just a little bit, of the statistical machine. Though this entails a steeper learning curve, the reward is a deeper understanding and greater flexibility in modeling. For example, once you have a posterior distribution (more on that soon) for the parameters in your model - you can derive a probability distribution for any quantity from those parameters. There are also situations where a Bayesian approach is the only feasible method. Finally, you can incorporate uncertainty from many sources in a logical, coherent manner (e.g., observation error, measurement error, variability due to random effects).