Hierarchical Bayesian Modeling

What is it?

Hierarchical models (also known as multilevel models or mixed effects models) are an extension of linear modeling which account for a hierarchical structure in the data such as ecosubsections within ecosections or even provinces. This thesis implements *Bayesian* hierarchical models to accomplish small area estimation on forest attributes, meaning that these models do not only take into account the hierarchical structure of the data, but they also incorporate prior information about the covariates in the model. To best understand a Bayesian hierarchical model, I will first begin with the standard OLS regression form. We have

$$Y_i = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon_i$$

where Y_i is the response variable for the ith observation and all observations are assumed to be independent. In this case, β_0 is the intercept term, ϵ_i is the unexplained error, and $x_1, \ldots, x_p$ are predictors or "covariates" with coefficient estimates $\beta_1, \ldots, \beta_p$ fit with the methods of Least Squares. The method of Least Squares estimates the intercept and slopes of these covariates by fitting a hyperplane which minimizes sum of squared residuals, where each residual is defined as $e_i = Y_{observed} - Y_{expected}$.

Bayesian hierarchical modeling does not use the method of Least Squares to estimate parameter values, instead we specify priors on each covariate and then sample from the posterior distribution using Markov Chain Monte Carlo (MCMC) methods. The most common MCMC method for hierarchical modeling is the Gibbs Sampler. THe Gibbs Sampler is useful because it uses the conditional distributions for each variable (which we will specify in our model) to estimate the joint distribution of our response variable.

Before we dig much deeper into how to use the Gibbs Sampler in this case, we should specify the model.

$$Y_{i,j} = X_i \vec{\beta}_i + Z_{i,j} b_{i,j} + \epsilon_{i,j}$$

Here $Y_{i,j}$ is the ith observation in the jth group, X_i is the ith element of the vector of fixed-effect predictors, that is, predictors that do not change depending on the group, these have coefficient estimates $\vec{\beta}_i$. Next, we have $Z_{i,j}$ which is the ith element of the jth group of the vector of random-effect predictors, that is, predictors that change depending on the group, these have coefficient estimates $b_{i,j}$. Finally, we have the unexplained error for the ith observation in the jth group, $\epsilon_{i,j}$.

Why use it?

observations do not need to be independent – we actually assume that they are not independent

"borrowing strength" / models that don't have amnesia

Example

think i'll be able to make this forestry related.