

# 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

$$\begin{aligned} Y_i &= \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p + \epsilon_i \\ &= \vec{\beta}_i X_i + \epsilon_i \end{aligned}$$

where  $Y_i$  is the response variable for the  $i$ th observation and all observations are assumed to be independent. In this case,  $\beta_0$  is the intercept term,  $\epsilon_i$  is the unexplained error, and  $x_1, \dots, x_p$  are predictors or “covariates” with coefficient estimates  $\beta_1, \dots, \beta_p$  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  $e_i = Y_{\text{observed}} - Y_{\text{expected}}$ .

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.

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

Here  $Y_{i,j}$  is the  $i$ th observation in the  $j$ th group,  $X_i$  is the  $i$ th 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  $i$ th element of the  $j$ th 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  $i$ th observation in the  $j$ th group,  $\epsilon_{i,j}$ .

## Why use it?

Now that we have explicitly written the model form out we can consider why one may want to use multilevel modeling in their research. I will focus on the main benefits of multilevel modeling to small area estimation. First, we have improved estimation for imbalance in sampling: multilevel models will automatically deal with issues of over- and under-sampling of clusters by assigning differing uncertainty across clusters. Also, when we are dealing with groups within the data, the multilevel models will model the variation between groups explicitly. Finally, with multilevel models we can avoid pre-averaging our data. This allows us to retain the variation in our data to be fully used in the model and we do not lose some valuable insight that this variation could give us. (Source: Statistical Rethinking, edition 2, page 414)

“borrowing strength” / models that don’t have amnesia

## Example

Now that we know what Hierarchical Bayesian model is, let’s think about using on in practice and the practical benefits we would gain from it in a concrete sense.