

Lecture 19—Wednesday, February 22, 2012

Topics

- Random effects approach to dealing with observational heterogeneity
- Multilevel formulation of the random slopes and intercepts model
- The likelihood of a mixed effects model
 - Likelihood of the data at the cluster level
 - Joint distribution of the data and random effects
 - Obtaining the marginal likelihood of the data
 - Contrasting fixed and random coefficients
- The random intercepts model: a minimal model for correlated data

Random effects approach to dealing with observational heterogeneity

The random effects analog of the separate slopes and intercepts regression model is the random slopes and intercepts model.

$$y_{ij} = \underbrace{\beta_0 + u_{0i}}_{\text{intercept}} + \underbrace{(\beta_1 + u_{1i})}_{\text{slope}} \text{Year}_{ij} + \varepsilon_{ij}$$

with $\varepsilon_{ij} \sim N(0, \sigma^2)$. In this model β_0 and β_1 represent the population-average coefficients while u_{0i} and u_{1i} are the deviations from this population average for coral core i . β_0 and β_1 can also be interpreted as the regression coefficients for a typical core, i.e., one corresponding to the middle of the distribution of random effects. The intercept for core i is $\beta_{0i} = \beta_0 + u_{0i}$ and the slope is $\beta_{1i} = \beta_1 + u_{1i}$. Thus the random slopes and intercepts formulation of this model is also the following.

$$y_{ij} = \beta_{0i} + \beta_{1i} \text{Year}_{ij} + \varepsilon_{ij}$$

What makes this model different from the fixed effects model is that u_{0i} and u_{1i} are not directly estimated but instead are assumed to be drawn from a multivariate normal distribution.

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right)$$

The diagonal entries of the multivariate normal covariance matrix are the individual variances of the intercept and slope random effects and the off-diagonal entry is their covariance. Because the correlation coefficient is defined by $\rho = \frac{\tau_{01}}{\tau_0 \tau_1}$, an equivalent way of writing this distribution (and the one used by R) is the following.

$$\begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \rho \tau_0 \tau_1 \\ \rho \tau_0 \tau_1 & \tau_1^2 \end{bmatrix} \right)$$

Rather than estimate the individual u_{0i} and u_{1i} we instead estimate the parameters of the covariance matrix of the multivariate normal distribution: ρ , τ_0 , and τ_1 . In the case of a normal response the likelihood simplifies in such a way that the individual u_{0i} and u_{1i} do not appear. For a non-normal response the random effects have to be numerically integrated out of the log-likelihood to yield what's called a marginal log-likelihood, which is then maximized.

The upstart is that in the random slopes and intercepts model we assume that the individual cores in the sample are drawn from a population of cores. The slopes and intercepts of individual cores are just typical random realizations from the population of slopes and intercepts from which they were drawn. Fig. 1 illustrates this situation for two specific cores that are numbered 1 and 2.

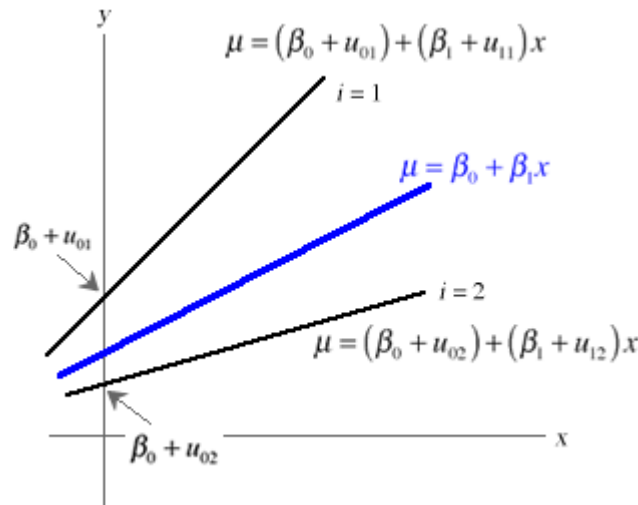


Fig. 1 Illustration of the random slopes and intercepts model

Multilevel formulation of the random slopes and intercepts model

It is common practice, particularly in sociology and political science, to formulate the random slopes and intercepts model slightly differently. Because it represents a hierarchy in the data, here annual rings nested in cores, it is convenient to write separate equations for each level of the hierarchy. These equations when combined yield the usual random slopes and intercepts model.

The level-1 model is a model for the individual rings contained within coral cores. Here, the level-1 model is just the trend model.

$$\text{Level 1 model: } y_{ij} = \beta_{0i} + \beta_{1i} \text{Year}_{ij} + \varepsilon_{ij}, \text{ where } \varepsilon_{ij} \sim N(0, \sigma^2), i = 1, \dots, g, j = 1, \dots, n_i$$

The individual rings are called the level-1 units while the cores are called the level-2 units. The subscripts on the intercept and slope indicate that these parameters can potentially vary across cores, i.e., each core has its own separate regression equation. These equations define the level-2 model.

$$\text{Level-2 Model: } \begin{aligned} \beta_{0i} &= \beta_0 + u_{0i} \\ \beta_{1i} &= \beta_1 + u_{1i} \end{aligned} \quad \text{where } \begin{bmatrix} u_{0i} \\ u_{1i} \end{bmatrix} \sim N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \tau_0^2 & \tau_{01} \\ \tau_{01} & \tau_1^2 \end{bmatrix} \right)$$

When we substitute the level-2 equations into the level-1 model we obtain the random slopes and intercepts model as before. In the multilevel formulation this combined expression is called the composite equation.

$$\text{Composite equation: } y_{ij} = \beta_0 + u_{0i} + (\beta_1 + u_{1i}) \text{Year}_{ij} + \varepsilon_{ij}$$

The multilevel formulation is a great bookkeeping tool. In the coral core model variables that characterize individual annual rings, such as Year, appear in the level-1 equation. Year is a level-1 variable. Variables that characterize individual cores (and therefore all of the rings contained in that core) appear in the level-2 equations. Such variables are called level-2 variables. Notice that we have a choice of whether we enter a level-2 variable into the level-2 intercept equation, the level-2 slope equation, or both.

As an example suppose we decide to add reef type as a predictor in our model. Reef type is a characteristic of cores and hence is a level-2 predictor and should be added to the level-2 equations. It is also a categorical variable with three levels and hence will have to be added to the regression equation as two dummy variables. In the equations below I add it to both the intercept and the slope equation thus allowing both the intercept and slope to vary depending on the type of reef from which the core came.

$$\text{Level 1 model: } y_{ij} = \beta_{0i} + \beta_{1i} \text{Year}_{ij} + \varepsilon_{ij}$$

Level-2 Model:

$$\beta_{0i} = \beta_0 + \gamma_1(\text{reef type} = 2) + \gamma_2(\text{reef type} = 3) + u_{0i}$$

$$\beta_{1i} = \beta_1 + \gamma_3(\text{reef type} = 2) + \gamma_4(\text{reef type} = 3) + u_{1i}$$

Now we have three population models, one for each reef type. The individual core regression lines in turn cluster about one of three population mean lines depending upon the reef type to which they belong. Fig. 2 illustrates this situation for two of the reef types, reef type 1 and reef type 2, and four cores, two from each reef type.

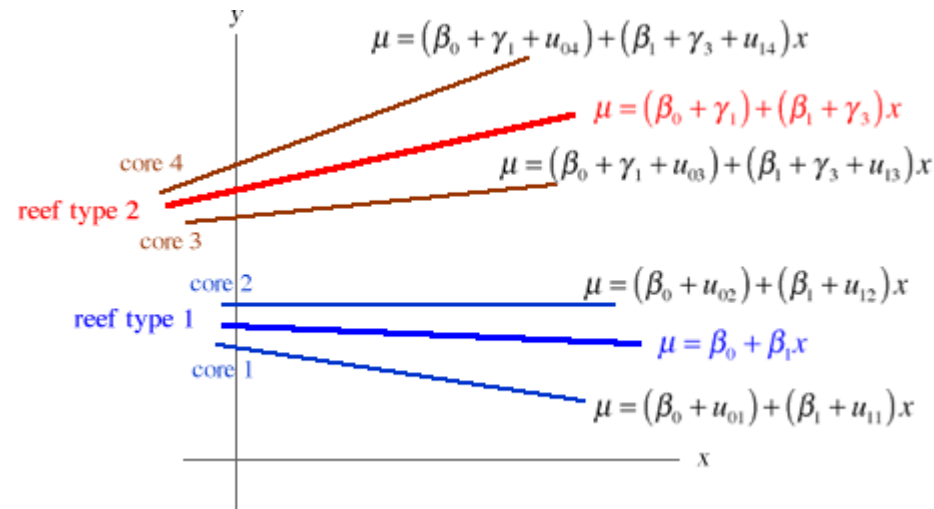


Fig. 2 Adding reef type as a level-2 predictor

From a modeling standpoint we would next test whether we need three population models (one for each reef type) to characterize the population of cores or just a single population model. Thus by using a mixed effects model we are able to include reef type in the model while at the same time letting each core have its own regression equation, something that was not possible in a purely fixed effects model.

The likelihood of a mixed effects model

The switch from a separate slopes and intercepts (pure fixed effects) model to a random slopes and intercepts model leads to a fundamental change in the likelihood that is being maximized. In what follows I use generic notation for the distributions involved because the ideas carry over to other distributions besides the normal distribution that we're focusing on at the moment. I assume that the observations come in clusters and that the clusters are a random sample from a population of clusters.

Likelihood of the data at the cluster level

Methods employing maximum likelihood all start from the same point, the likelihood function. Recall that if we have a random sample of observations, y_1, y_2, \dots, y_m , that are independent and identically distributed with probability density (mass) function $p(y_i, \theta)$ where θ is potentially a vector of parameters of interest, then the likelihood of our sample is the product of the individual density functions. For the normal regression model θ will consist of the set of regression parameters β and the variance of the response σ^2 , so that the likelihood can be written as follows.

$$p(\mathbf{y} | \beta, \sigma^2) = \prod_{i=1}^m p(y_i | \beta, \sigma^2)$$

With correlated data, i.e., structured data, the situation is more complicated. The data consist of observations y_{ij} where i indicates the level-2 unit and j the level-1 observation within that unit. Typically the level-2 units will be independent but the level-1 units will not be. If we let

$$\mathbf{y} = [y_{11}, y_{12}, \dots, y_{1n_1}, y_{21}, y_{22}, \dots, y_{2n_2}, \dots, y_{m1}, y_{m2}, \dots, y_{mn_m}]$$

so that $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{in_i}]$, then we can write the likelihood in terms of the m clusters as follows where I include additional parameters that relate to the distribution of clusters.

$$p(\mathbf{y} | \beta, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) = \prod_{i=1}^m p(\mathbf{y}_i | \beta, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01})$$

Joint distribution of the data and random effects

If we choose to model the cluster structure using a mixed effects model then the y_{ij} are not the only random quantities to consider. The intercept u_{0i} and slope u_{1i} random effects also have a distribution. Thus the natural starting place in constructing the likelihood is with the joint density function of the data and random effects in cluster i , $f(y_{i1}, y_{i2}, \dots, y_{in_i}, u_{0i}, u_{1i})$. From elementary probability theory we can write

$$f(y_{i1}, y_{i2}, \dots, y_{in_i}, u_{0i}, u_{1i}) = g(y_{i1}, y_{i2}, \dots, y_{in_i} | u_{0i}, u_{1i}) h(u_{0i}, u_{1i})$$

- Here $g(y_{i1}, y_{i2}, \dots, y_{in_i} | u_{0i}, u_{1i})$ is the conditional distribution of the observed data assuming the values of the random effects are known. This will correspond to the probability model we assume for the data at level 1. For normal models this density will also be a function of the

vector of fixed effect parameters $\boldsymbol{\beta}$ and the variance σ^2 .

- $h(u_{0i}, u_{1i})$ is the joint distribution of the random effects. In all of the examples in this course this will be assumed to be multivariate normal with mean zero and variance matrix that was given above.

Let $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{in_i}]$. If we include the regression parameters in the likelihood then the joint distribution of the data and the random effects in cluster i would be written as follows

$$f(\mathbf{y}_i, u_{0i}, u_{1i} | \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) = g(\mathbf{y}_i | u_{0i}, u_{1i}, \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) h(u_{0i}, u_{1i} | \tau_0^2, \tau_1^2, \tau_{01})$$

Conditional on the value of the random effects, the individual y_{ij} in cluster i are assumed to be independent. So we have

$$\begin{aligned} f(\mathbf{y}_i, u_{0i}, u_{1i} | \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) &= g(\mathbf{y}_i | u_{0i}, u_{1i}, \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) h(u_{0i}, u_{1i} | \tau_0^2, \tau_1^2, \tau_{01}) \\ &= \prod_{j=1}^{n_i} k(y_{ij} | u_{0i}, u_{1i}, \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) h(u_{0i}, u_{1i} | \tau_0^2, \tau_1^2, \tau_{01}) \end{aligned}$$

Here I use $k(y_{ij} | u_i, \boldsymbol{\beta}, \sigma^2, \tau^2)$ to denote the distribution of the individual y_{ij} conditional on the random effects and the regression and variance parameters.

Obtaining the marginal likelihood of the data

To go from a joint probability of y_{ij} , u_{0i} , and u_{1i} to a marginal probability for y_{ij} we integrate over the possible values of the random effects. Thus the marginal likelihood for cluster i is the following.

$$\begin{aligned} p(\mathbf{y}_i | \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) &= \int f(\mathbf{y}_i, u_{0i}, u_{1i} | \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) d\mathbf{u}_i \\ &= \int g(\mathbf{y}_i | u_{0i}, u_{1i}, \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) h(u_{0i}, u_{1i} | \tau_0^2, \tau_1^2, \tau_{01}) d\mathbf{u}_i \end{aligned}$$

Putting all these pieces together we can write the marginal likelihood of our data as follows.

$$\begin{aligned}
p(\mathbf{y}|\boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) &= \prod_{i=1}^m p(\mathbf{y}_i|\boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) \\
&= \prod_{i=1}^m \int f(\mathbf{y}_i, u_{0i}, u_{1i}|\boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) d\mathbf{u}_i \\
&= \prod_{i=1}^m \int g(\mathbf{y}_i|u_{0i}, u_{1i}, \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) h(u_{0i}, u_{1i}|\tau_0^2, \tau_1^2, \tau_{01}) d\mathbf{u}_i \\
&= \prod_{i=1}^m \int \prod_{j=1}^{n_i} k(y_{ij}|u_{0i}, u_{1i}, \boldsymbol{\beta}, \sigma^2, \tau_0^2, \tau_1^2, \tau_{01}) h(u_{0i}, u_{1i}|\tau_0^2, \tau_1^2, \tau_{01}) d\mathbf{u}_i
\end{aligned}$$

For the random intercepts model we've been considering the distribution of the random effects is normal and the distribution of the response given the random effects is normal. Thus both functions k and h in the last expression are normal densities. It turns out under this scenario that p is also a normal density and its mean and variance can be found from the means and variances of k and h . Thus with a normal model it is possible to find the marginal likelihood without doing the integration shown in the last expression. This is what the **lme** function of the **nlme** package does. For models in which k is not a normal distribution, e.g., binomial, Poisson or some other member of the exponential family, the integration needs to be carried out numerically in order to obtain the marginal likelihood. This is what the **lmer** function of the **lme4** package does.

Contrasting fixed and random coefficients

It is worth comparing what the likelihood looks like for the separate slopes and intercepts model.

$$L(\mathbf{y}|\boldsymbol{\beta}, \sigma^2) = \prod_{i=1}^m \prod_{j=1}^{n_i} \ell(y_{ij}|\boldsymbol{\beta}, \sigma^2)$$

In this likelihood the vector $\boldsymbol{\beta}$ will also include the coefficients of the various dummy variables (now fixed effects) that are used to indicate group membership. These can then be used to obtain estimates of the slopes and intercepts of each group. So in the separate intercepts model we obtain estimates of all the group slopes and intercepts. In the random intercepts model we assume the intercepts arise from a common distribution and then estimate only the parameters of that distribution. If so then how can we get predictions of the intercept and slope random effects. How is this possible if the random effects are integrated out of the likelihood and are no longer present?

The predictions of the random effects are not obtained as a result of maximizing the likelihood. What happens is that after having estimated the regression parameters and variances in the marginal likelihood these estimates are used to obtain predictions of the random effects. Formally the predictions are the conditional means of the random effects distribution given the data and the regression parameters and are obtained using what amounts to a generalized least squares formula. The fact that they are the means of distributions gives them a distinct Bayesian flavor and they're

sometimes called empirical Bayes estimates (also best linear unbiased predictors or BLUPs). So, unlike the separate slopes and intercepts model the steps taken to obtain predictions of the random slopes and intercepts are entirely post hoc.

The random intercepts model: a minimal model for correlated data

Fitting a random intercepts model is a minimal way to account for lack of independence in a hierarchical experimental design. Suppose we have a split plot design in which treatment A with two levels is randomly applied to groups (whole plots) and treatment B with two levels is random applied to individuals (split plots) separately within each group. The whole plots could be aquariums (with whole plot treatment "temperature") and the split plots could be individual fish in the tanks (with split plot treatment "species of fish").

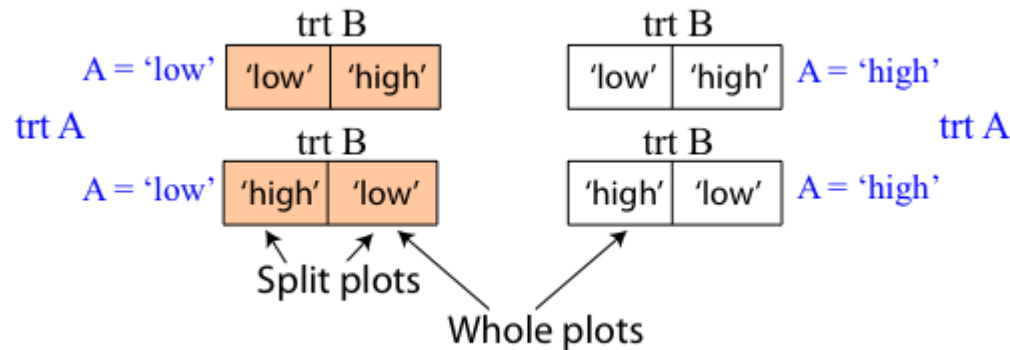


Fig. 3 Split plot design with whole plot treatment A and split plot treatment B

A minimal additive model that addresses the observational heterogeneity present in this design is the following random intercepts model (written in composite form).

$$y_{ij} = \beta_0 + \beta_1 \text{trtA}_i + \beta_2 \text{trtB}_{ij} + u_{0i} + \varepsilon_{ij}$$

where $u_{0i} \sim N(0, \tau^2)$ and $\varepsilon_{ij} \sim N(0, \sigma^2)$ and **trtA** and **trtB** are dummy variables indicating the treatment. (A more general model would allow also for an interaction between A and B but this is an unnecessary complication for the point I wish to make.)

Consider the following three experimental units: two of them come from same group (level-2 unit r) and one comes from a different group (level-2 unit s). Here are their three regression equations.

$$\left. \begin{aligned} y_{r1} &= \beta_0 + \beta_1 \text{trtA}_r + \beta_2 \text{trtB}_{r1} + \textcolor{red}{u}_{0r} + \varepsilon_{r1} \\ y_{r2} &= \beta_0 + \beta_1 \text{trtA}_r + \beta_2 \text{trtB}_{r2} + \textcolor{red}{u}_{0r} + \varepsilon_{r2} \end{aligned} \right\} \text{Level-2 unit } r$$

$$y_{s3} = \beta_0 + \beta_1 \text{trtA}_s + \beta_2 \text{trtB}_{s3} + \textcolor{blue}{u}_{0s} + \varepsilon_{s3} \} \text{Level-2 unit } s$$

Notice that both observations from level-2 unit r share the common random effect u_{0r} . The observation from level-2 unit s on the other hand possesses a different random effect, u_{0s} . The random effect u_{0r} then accounts for all those unmeasured characteristics that make group r different from any other group. Its presence causes members of this group to be similar to each other but different from members of other groups.

It can be shown that the random intercepts model induces a correlation in the observations coming from the same group. The relevant covariances are the following.

$$\text{Cov}(Y_{ij}, Y_{kl}) = \begin{cases} 0, & \text{if } i \neq k & \text{different groups} \\ \tau^2, & \text{if } i = k \text{ and } j \neq l & \text{same group, different observations} \\ \tau^2 + \sigma^2, & \text{if } i = k \text{ and } j = l & \text{same group, same observation} \end{cases}$$

Observations from the same group ($i = k$) have a positive covariance, τ^2 , while observations from different groups ($i \neq k$) have a covariance of zero. So observations coming from the same group are more similar (have positive covariance) than observations coming from different groups (zero covariance).

Using the definition of correlation we can translate this statement about covariance into one about correlation. For distinct observations with $j \neq l$ we have

$$\begin{aligned} \rho(Y_{ij}, Y_{kl}) &= \frac{\text{Cov}(Y_{ij}, Y_{kl})}{\sqrt{\text{Var}(Y_{ij})} \sqrt{\text{Var}(Y_{kl})}} = \frac{\text{Cov}(Y_{ij}, Y_{kl})}{\sqrt{\text{Cov}(Y_{ij}, Y_{ij})} \sqrt{\text{Cov}(Y_{kl}, Y_{kl})}} \\ &= \begin{cases} \frac{\tau^2}{\tau^2 + \sigma^2}, & \text{if } i = k \\ 0, & \text{if } i \neq k \end{cases} \end{aligned}$$

and so we see that observations coming from the same level-2 unit (group) are correlated while observations from different level-2 units are uncorrelated.

Observe that the correlation structure is really very simple. Any two observations from the same level-2 unit have the same correlation. If we let

$\phi = \frac{\tau^2}{\tau^2 + \sigma^2}$ then the correlation matrix for any level-2 unit takes the following form.

$$\rho = \begin{bmatrix} 1 & \phi & \cdots & \phi \\ \phi & 1 & & \vdots \\ \vdots & & \ddots & \phi \\ \phi & \cdots & \phi & 1 \end{bmatrix}$$

Such a constant correlation matrix is called exchangeable. If we assemble these exchangeable correlation matrices ρ for all the level-2 units into a single matrix in which all observations are represented we obtain the following block-diagonal correlation structure for the model.

$$\mathbf{R} = \begin{bmatrix} \rho_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \rho_2 & & \vdots \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \rho_n \end{bmatrix}$$

where each ρ_i is an exchangeable correlation matrix like the one shown above and $\mathbf{0}$ is a matrix of zeros.

[Course Home Page](#)

Jack Weiss

Phone: (919) 962-5930

E-Mail: jack_weiss@unc.edu

Address: Curriculum for the Environment and Ecology, Box 3275, University of North Carolina, Chapel Hill, 27599

Copyright © 2012

Last Revised--February 22, 2012

URL: https://sakai.unc.edu/access/content/group/2842013b-58f5-4453-aa8d-3e01bacbfc3d/public/Ecol562_Spring2012/docs/lectures/lecture19.htm