FISH 559: Numerical Computing for the Natural Resources
Homework 1 : Mixed Effect Models (out of 105 points)

Question 1 (40 points)
Implement a linear mixed effects model numerically using R to replicate the maximum likelihood estimates of the population mean density, the extent of among-population variation and the extent of within-population variation using the Streams data (Home1a.txt). Compute a likelihood profile for the population mean density. Compare the 95% confidence intervals from this likelihood profile with that determined using LME.

Question 2 (65 points)
Dorn (2002; *North American Journal of Fisheries Management* 22: 280-300) used a Bayesian mixed effects model to develop a Bayesian prior distribution for the steepness[1] of the stock-recruitment relationship for U.S. West Coast rockfish species. The Ricker stock-recruitment relationship can be reparameterized as follows:

$$R_y = \frac{\alpha}{S\tilde{S}B} S_y \, e^{-\beta S_y}$$

where R_y is the recruitment for year *y*,
S_y is spawning stock biomass (or a proxy thereof) for year *y*,
$S\tilde{S}B$ is the spawner biomass-per-recruit in the absence of exploitation, and
α, β are the parameters of the stock-recruitment relationship.

Use R to fit a linear random effects model to data for 11 West Coast rockfish species treating $\ell n\alpha$ as a random effect and β as a (species-specific) fixed effect. Assume that the noise about the stock-recruitment relationship is log-normal and assume that the variation about the stock-recruitment relationship is the same for all species.

The data are stored in the file Home1b.txt (format: Spawning stock biomass, recruitment, species code, spawner biomass-per-recruit).

Evaluate the fits using plots of residuals, random effects, etc.

There are *N* streams and each streams has n_i observations. Let $y_{i,k}$ be the observation for the *k*'th site in steam *i* and **D** denote the entire data set. The parameters of the model are the mean (β), the between-stream variation in average density (σ_b), the residual

[1] Steepness is the fraction of the recruitment at the virgin spawning stock biomass expected when the spawning stock biomass is reduced to 20% of its unfished level.

variation (σ_ε), and the random effects $\{b_i : i = 1, 2, .., N\}$. The likelihood for a random effects model integrates over the random effects, i.e.:

$$L(\mathbf{D} \mid \beta, \sigma_b, \sigma_\varepsilon) = \int L(\mathbf{D} \mid \beta, \sigma_b, \sigma_\varepsilon, \underline{b})\,\mathrm{P}(\underline{b} \mid \sigma_b)\,\mathrm{d}\underline{b} \tag{1}$$

Now, each data point is associated with only one random effect so we can simplify the integral from one six dimension integral to six one dimensional integral (see discussion of why we should do this in lecture 2), i.e.:

$$L(\mathbf{D} \mid \beta, \sigma_b, \sigma_\varepsilon) = \prod_i \int L(\mathrm{D}_i \mid \beta, \sigma_b, \sigma_\varepsilon, b_i)\,\mathrm{P}(b_i \mid \sigma_b)\,\mathrm{d}b_i \tag{2}$$

where D_i is the data set for stream *i* and b_i is the random effect for stream *i*. Now, the likelihood of each data point is normal so:

$$L(\mathrm{D}_i \mid \beta, \sigma_b, \sigma_\varepsilon, b_i) = \prod_k \tfrac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{(y_{i,k}-\beta-b_i)^2}{2\sigma_\varepsilon^2}} \tag{3}$$

Given that $\mathrm{P}(b_i \mid \sigma_b)$ is also normal, we can write Equation 2 as:

$$L(\mathbf{D} \mid \beta, \sigma_b, \sigma_\varepsilon) = \prod_i \int \tfrac{1}{\sqrt{2\pi}\sigma_b} e^{-\frac{b_i^2}{2\sigma_b^2}} \prod_k \tfrac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{(y_{i,k}-\beta-b_i)^2}{2\sigma_\varepsilon^2}}\,\mathrm{d}b_i \tag{4}$$

You need to write a function to do each of the N integrations included in the RHS of this equation, a function which computes Equation 4 using that function and code to use the mle function in R to find the best estimates for β, σ_b, and σ_ε.

Two hints:

1. The integral in Equation 4 is from $-\infty$ and ∞ and if you use Simpsons rule (see the equation at the bottom of slide 10 of Friday's lecture) you will need many many function calls (trust me). However, a trick is to scale the b_i by σ_b so that Equation 4 becomes:

$$L(\mathbf{D} \mid \beta, \sigma_b, \sigma_\varepsilon) = \prod_i \int \tfrac{1}{\sqrt{2\pi}} e^{-\frac{b_i^2}{2}} \prod_k \tfrac{1}{\sqrt{2\pi}\sigma_\varepsilon} e^{-\frac{(y_{i,k}-\beta-b_i\sigma_b)^2}{2\sigma_\varepsilon^2}}\,\mathrm{d}b_i \tag{5}$$

   You can now get away with a range for b_i from about -5 to 5 (but chose a small step size anyway).

2. Run the streams data set through lme (with the method=ML option) to get the correct values for the parameters to check your answers.