Advanced Modeling in R

Non-linear, Bayesian, and mixed effect methods

R. Condit, M. Ferrari
Cenpat, Patagonia
October 2012

Assignments

Tuesday AM, 9 October

1. **Regression: Biomass data**

   Use log-transformed data
   Possible terms
   a) simple model: log(AGB) vs. log(dbh)
   b) second order term: log(dbh) squared
   c) rainfall and elevation

   Graph
   a) log(AGB) vs. log(dbh)
   b) add curve of best fit
   c) overlay curves for high rainfall and low rainfall

   Character variable (factors)
   a) use ForestType in the model
      - $\log(volume) \sim \log(dbh) + ForestType$
   b) compare to 3 independent models (3 forest types)
      - $\log(volume[dry]) \sim \log(dbh[dry])$
      - etc.

Tuesday PM, 9 October

1. **Fit a linear model with variable SD**

   - Data
     a) pupsize: Wt as a function of momcat

   b) cecrin: growth (gr12) as a function of diameter (dbh1)
   c) treemass: log(agb) as a function of log(dbh)
- Likelihood function is provided (llike.linearmodel.full, in teaching.functions.r)
- Requires 4 parameters: slope, intercept of model, then slope, intercept for SD

2. **Fit a non-linear model to quantitative data**
   - Model types
     a) $y = H\left(1 - e^{-ax^b}\right)$ (it's in teaching.functions.r)
     b) $y \sim x + \log x$ (write yourself)
   - Data
     a) pupsize: Wt is a non-linear function of momage
     b) treeht: ht is a non-linear function of dbh (extract one species, eg quaras or tri2tu or pri2co)
   - Assume error is Gaussian (dnorm)

3. **Save a graph of one or more of the model fits above and email to me (conditr@gmail.com)**

Wednesday AM, 10 Octuber

1. **Using llike.generalmodel, write and test a quadratic function, $a + bx + cx^2$, with any of the data**

2. **Fit a two-parameter survival model**
   - Data: cecrin
   - $logit(status3) \sim gr12 + dbh2$ (write yourself; see likelihood functions in teaching.functions.r for help)

3. **Simulate regression with error in x**
   a) Define a variable xerr as the standard deviation due to error
   b) Use rnorm to generate xobs (contrast with xtrue)
   c) Determine how estimated regression coefficient is affected

4. **Simulate multi-level regression where slope varies between groups**
   a) Define a variable slopesd defining standard deviation of slope across groups
   b) Use rnorm to simulate values of slope in three groups
   c) Merge with rbind into one dataframe
   d) Extra: write a loop to simulate a larger number of groups
   e) Use lm for regression, $y \sim x + group$
   f) Use lmer for model, $y \sim 1 + x + (1 + x|group)$

5. **Extra credit: Simulate regression with two predictors, x_1 and x_2**
   a) Define a correlation between x_1 and x_2
   b) Use rnorm to generate x_2 then y

c) Determine how estimated regression result is affected

6. **Extra credit: Simulate Poisson regression**
   a) Define a variable xerr as the standard deviation due to error
   b) Use rpois to generate xobs (contrast with xtrue)

Thursday, 11 October

1. **Write a function to create 100 different populations with known size (N), known means, 2 levels of SD, normal distribution**
   a) Create a separate file to hold the function, to be sourced as updated, and with comments
   b) Requires a loop of 100
      i. Each step create one sample using the within-group SD
      ii. Start with one example where within-group means are identical
      iii. Then allow the group means to vary following an overarching normal distribution (the hyper-distribution)
   c) Save results into one table
      i. Table has 100 rows for 100 populations
      ii. Each row has 5 columns: N for population size, the known mean and SD, and the observed mean and SD
   d) Repeat with different N's (large or small sample) and varying the hyper-SD
   e) Can you figure out how to repeat so the N's vary?
   f) Send me the file with function(s) via email

Friday, 12 October

1. **Use lmer for regression of logagb on logdbh with species and forest type as factors**
   a) Include squared term for logdbh
   b) Add locality as a group effect (does it change the fixed effect)
   c) Test forest type as fixed effect and as group effect
   d) Graph points and lines
   e) Compare alternative models

2. **Use lmer for regression of pup Wt on momcat with year as a mixed effect**
   a) Variable intercept, slope, or both
   b) Graph all points
   c) Use xyplot for groups (Lattice)
   d) Overlay lines of all random effects
   e) Compare alternative models

3. **Use lmer for regression of log(ht) on log(dbh) with species as a mixed effect**

Saturday, 13 October

1. **Fit a linear model to logagb vs log using the Bayesian method (modelfitBayes.r)**

2. **Use the function linearModelHier in growthModel.HierBayes to fit a hierarchical linear model**

   a) Locality as the random (group) factor: use dataset treemass.split
   b) Graph a line for each locality
   c) The model will also work for growth data, using dataset BCIlist, but more slowly

3. **Test the function survivalModel.Gibbs in survivalGibbs.r to fit a survival model (it's finished and should work)**