Regression and Modeling in R

R. Condit
Universidade de São Paulo, IB-Depto Ecologia
24-28 October 2011
Assignments

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.

2. **Regression: Function to calculate linear regression**

   The basics

   a) Accept x and y values
   b) Return regression parameters
   c) Graph y vs. x
   d) Adds regression line

   More information

   a) Find predicted y for every x
   b) Find residuals and their variance
   c) Graph residuals

   Advanced (extra credit)

   a) Include option of logarithmic transformation
   b) Allow multiple x predictors
   c) Graph y as a function of a chosen x

3. **Simulation: Create a simulated correlation and test how well lm fits the parameters**

   The basics

   a) Define x from normal distribution
   b) Define slope and intercept parameters
   c) Define error with *rnorm* and *sd*
   d) Calculate y
   e) Use *lm* to estimate slope and intercept

   More information

   a) Evaluate impact of increasing error
   b) Evaluate impact of error in measuring x
   c) Evaluate impact of highly non-Gaussian x

   Advanced (extra credit)

   a) Test multiple regression, with x_1 and x_2 predictors
   b) Evaluate impact of correlation between x_1 and x_2

4. **Non-linear model**
   - Use optim to fit a model $y \sim x + x^2$, or $y \sim x + \log x$
   - Fit a model to tree height data (dataset treeht)
   - Consider this function of height as a function of dbh:
   - $h = H_{max}\left(1 - e^{-ax^b}\right)$ where H_{max}, a, and b are parameters, h is height and x is dbh.
   - Use data from a single species to estimate the 3 parameters using maximum likelihood and a Gaussian error.

5. **Error functions**
   - dlnorm
     - for abundances, whether integer or not
     - good match for tree growth rates
     - but cannot handle zeroes
   - dpois including zeroes (but does not handle most ecological data well)
     - for integer abundances
     - handles zeroes
     - however, close to Gaussian so not appropriate for much ecological data
   - dnbinom (common in ecology)
     - for integer abundances that are highly skewed
     - very common in ecology
     - R: prob=dnbinom(count,size=k,mu=mu)
     - size is 'clumping parameter'; mu is mean

6. **Two-parameter survival model**
    - Fit tree survival to growth and dbh
    - $logit(status_2) \sim gr_1 + dbh_1$

7. **Program a Gibbs sampler for tree height model**
   - Write the sampler for the 3-parameter height model (formula in assignment 3)
   - Extend the program to allow a linear $height \sim dbh$ model
   - I provide a head start (*modelGibbs.r*)