An Introduction to Mixed-effects Models

Andrew Robinson

Department of Mathematics & Statistics
University of Melbourne

February 17, 2006

Outline

1. What are they?
2. Why use them?
3. Example
   - The Data
   - Analysis
4. Thinking about Effects
5. A Modelling Strategy

Characteristics

Mixed-effects models incorporate *two kinds* of predictor variables.

- Fixed effects - speak for themselves.
- Random effects - represent a population.

Necessity

Natural resources data commonly have hierarchical structure.

- Trees within plots within stands within forests.
- Times within trees ...

Mixed-effects models enable the modeling of correlated data *without* violation of important regression assumptions.

Necessity

Natural resources data commonly have hierarchical structure.

- Trees within plots within stands within forests.
- Times within trees . . .

Mixed-effects models enable the modeling of correlated data *without* violation of important regression assumptions.

Regression assumptions.

- True relationship is linear.
- Residuals are normally distributed.
- Residuals have identical distribution (variance).
- Residuals are independent.

Utility

Mixed effects models allow the estimation of useful quantities.

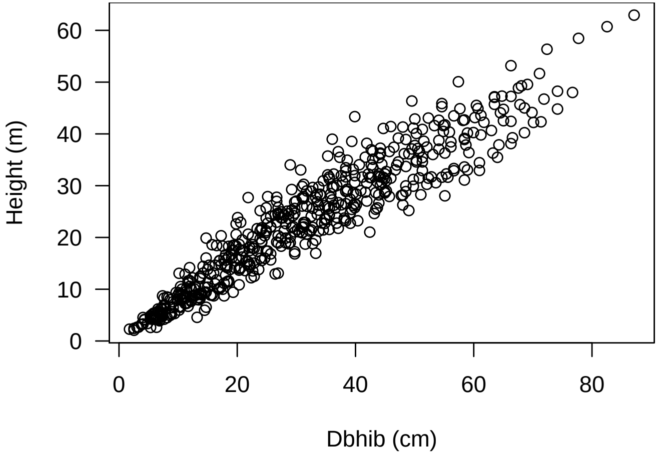
- Variance components.
- Intra-class correlation.

Data - Height/Diameter from Stage (1963)

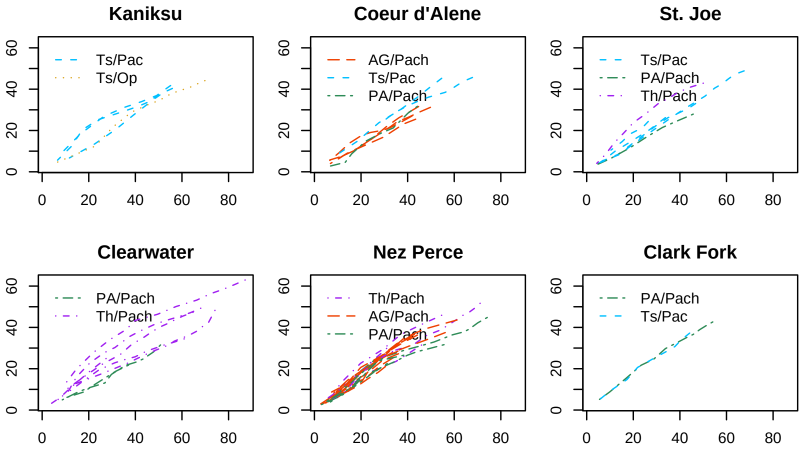
A brief synopsis: a sample of 66 trees was selected in national forests around northern and central Idaho. According to Stage (*pers. comm.* 2003), the trees were selected purposively.

The habitat type and diameter at 4'6" were also recorded for each tree, as was the national forest from which it came. Each tree was then split, and decadal measures were made of height and diameter inside bark at 4'6".

Scatterplot



Another look



Model for getting it wrong in R

$$h_i = \beta_0 + \beta_1 \times d_i + \epsilon_i \tag{1}$$

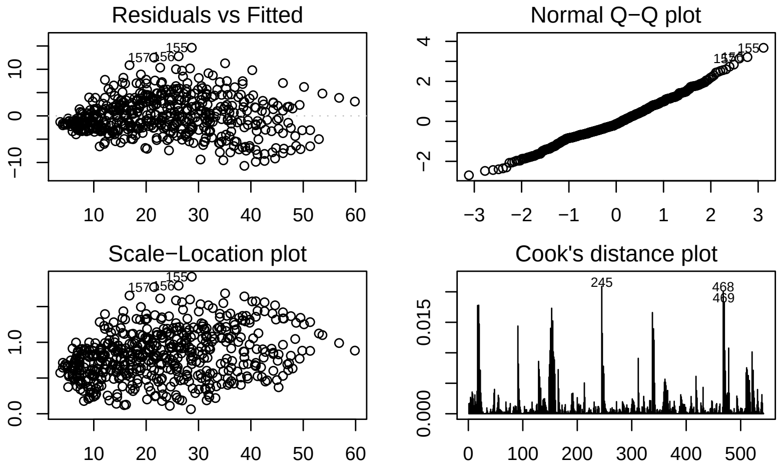
Model for getting it wrong in R

$$h_i = \beta_0 + \beta_1 \times d_i + \epsilon_i \tag{1}$$

Regression assumptions.

- True relationship is linear.
- $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$
- $Cov(\epsilon_i, \epsilon_j) = 0$ for $i \neq j$

Diagnostics for getting it wrong in R



Model for getting it less wrong in R

$$h_{it} = \beta_0 + (\beta_1 + b_{1i}) \times d_{it} + \epsilon_{it} \tag{2}$$

Model for getting it less wrong in R

$$h_{it} = \beta_0 + (\beta_1 + b_{1i}) \times d_{it} + \epsilon_{it} \tag{2}$$

Regression assumptions.

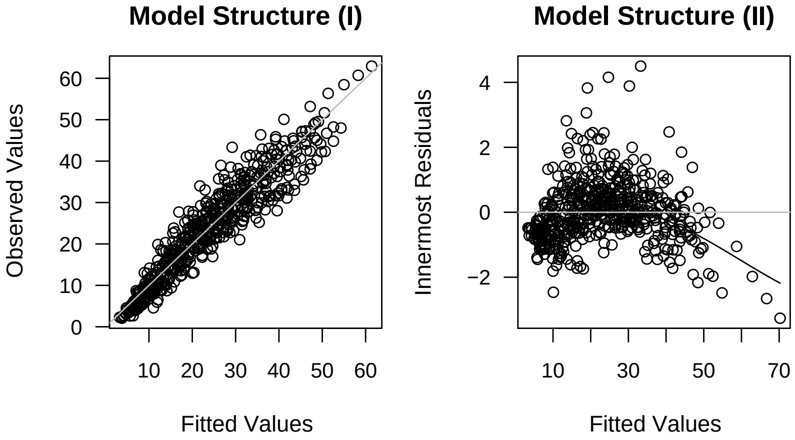
- True relationship is linear.
- $b_{1i} \sim \mathcal{N}(0, \sigma^2_{b_1})$
- $\epsilon_{it} \sim \mathcal{N}(0, \sigma^2)$
- $Cov(\epsilon_{it}, \epsilon_{jt}) = 0$ for $i \neq j$
- $Cov(\epsilon_{it}, \epsilon_{ig}) = 0$ for $t \neq g$

Assumptions for getting it less wrong in R

Now, the key assumptions that we're making are that:

1. the model structure is correctly specified
2. the tree and forest random effects are normally distributed,
3. the tree random effects are homoscedastic within the forest random effects.
4. the inner-most residuals are normally distributed,
5. the inner-most residuals are homoscedastic within and across the tree random effects.
6. the innermost residuals are independent within the groups.

Diagnostics for getting it less wrong in R



Diagnostics for getting it less wrong in R

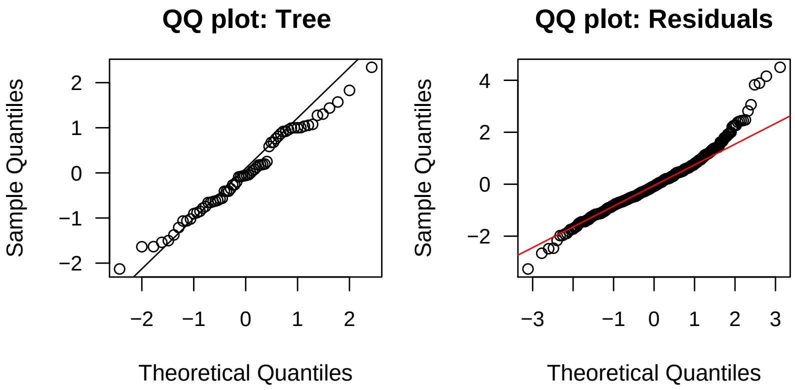
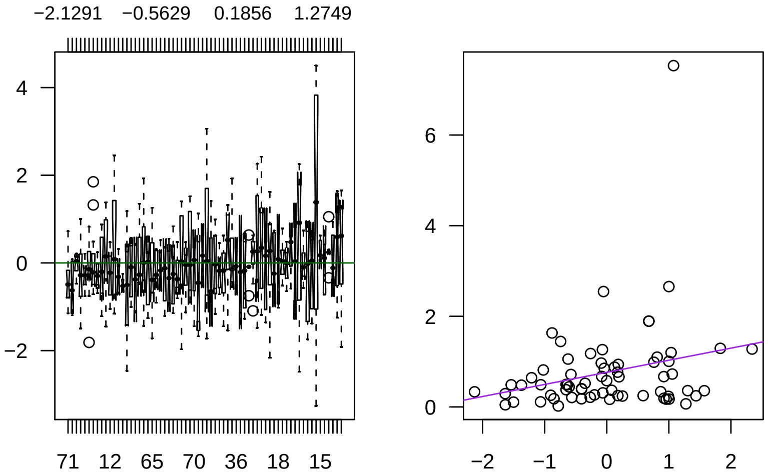
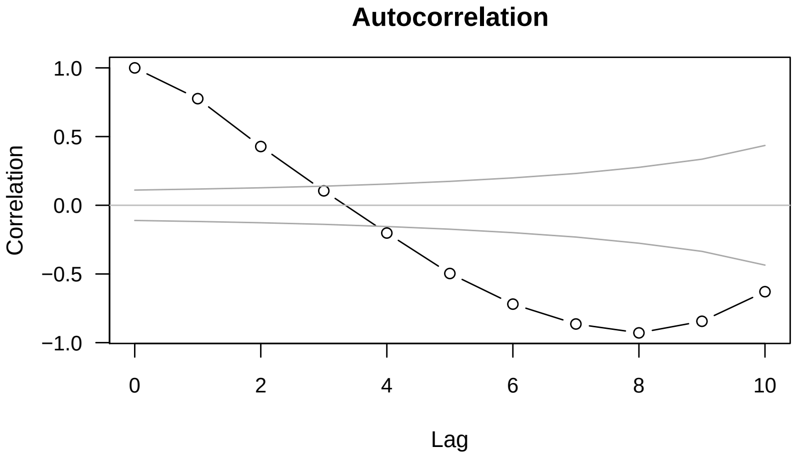


Figure: More useful regression diagnostics from R.

Diagnostics for getting it less wrong in R



Diagnostics for getting it less wrong in R



The roles differ

For the design,

The roles differ

For the design,

- fixed effects represent *themselves*;
- random effects represent *a population*.

The roles differ

Within the model,

The roles differ

Within the model,

- fixed effects *explain* variation;
- random effects *organize* unexplained variation.

Another perspective

Random effects are effects that common sense says will explain variation, but you don't want to have to know them in order to be able to apply the model.

Modelling is much more involved

Add a new dimension to your flow chart!

A Modelling Strategy

The modeling strategy depends on the modelers intention.

1. Fit baseline model.
   1. Include the meaningful fixed effects.
   2. Include the design random effects.
2. Check the assumption diagnostics.
3. Add or modify random components until diagnostics are satisfied.
   1. a heteroskedastic variance structure (several candidates)
   2. a correlation structure (several candidates)
   3. extra random effects (e.g. random slopes)
4. Consider adding more fixed effects.
5. Re-examine the diagnostics, add/modify random effects, etc.