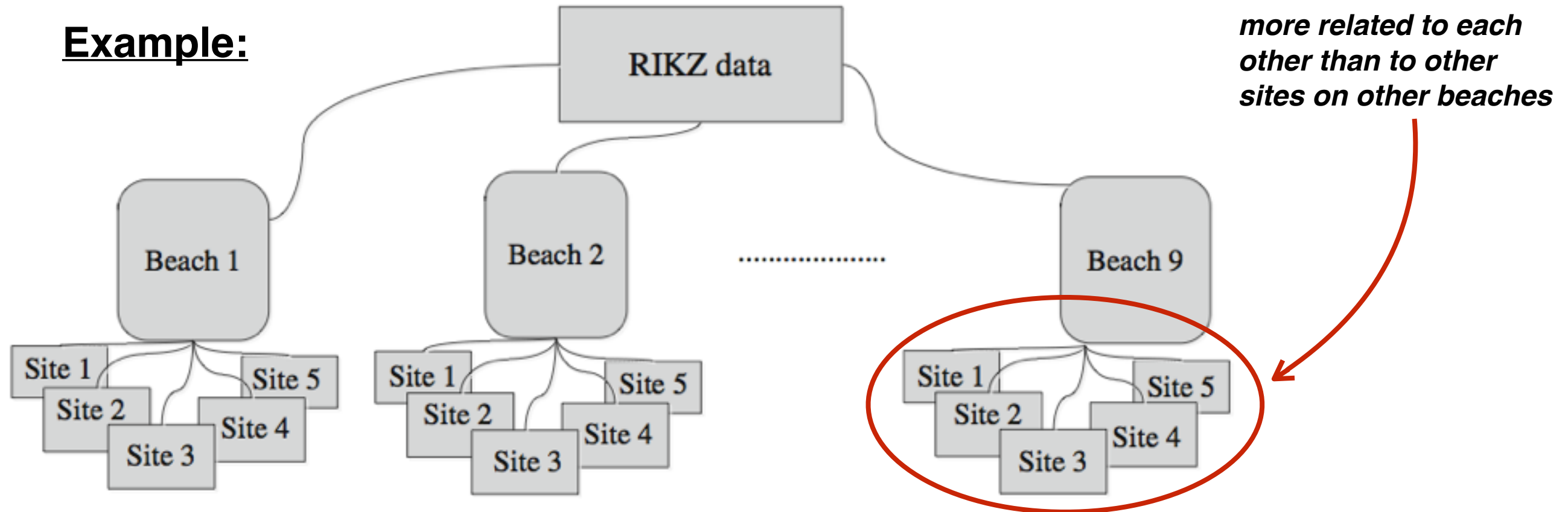


RClub2

Ch. 5.1-5.3

What is a multilevel / mixed / nested model?

Example:



DV: Species Richness (R)

Predictors: Exposure and NAP (height above sea level)

Expt: Measured DV/predictors at 9 beaches (at 5 sites within a beach)

Is there a relationship between species richness and NAP?

$$R_{ij} = \alpha + \beta_1 \times NAP_{ij} + \beta_2 \times Exposure_i + \varepsilon_{ij} \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

Examples of hierarchical/nested designs:

- Children with the same parents tend to be more alike in their physical and mental characteristics than individuals chosen at random from the population at large.
- Individuals may be further nested within geographical areas or institutions such as schools or employers.
- Within-subjects designs with repeated/longitudinal studies where an individual's responses over time are correlated with each other.

Why use multilevel models?

- observations are not independent, therefore standard errors of regression coefficients will be underestimated, leading to overestimation of significance
- interest in group effects vs individual differences (i.e. exploring the error term)
- possible to separate out effects due to observed and unobserved group characteristics
- inference to a *population* (i.e. generalizability of results)

2 Stage Analysis

Step 1: Run model for each beach

$$R_{ij} = \alpha + \beta_i \times NAP_{ij} + \varepsilon_{ij} \quad j = 1, \dots, 5 \quad (5.2)$$

This process is then carried out for data of each beach in turn. In a more abstract matrix notation, we can write the model for the data of beach i as

$$\begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ R_{i4} \\ R_{i5} \end{pmatrix} = \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i1} \\ 1 & NAP_{i1} \\ 1 & NAP_{i1} \\ 1 & NAP_{i1} \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix} \Leftrightarrow \mathbf{R}_i = \mathbf{Z}_i \times \boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i \quad (5.3)$$

\mathbf{R}_i is now a vector of length 5 containing the species richness values of the 5 sites on beach i .

Step 2: model estimated regression coefficients as a function of exposure

$$\hat{\beta}_i = \eta + \tau \times Exposure_i + b_i \quad i = 1, \dots, 9 \quad (5.4)$$

This is ‘just’ a one-way ANOVA. The response variable is the estimated slopes from step 1, Exposure is the (nominal) explanatory variable, τ is the corresponding regression parameter, η is the intercept, and b_i is random noise. The matrix notation

$$\begin{pmatrix} -0.37 \\ -4.17 \\ -1.75 \\ -1.24 \\ -8.90 \\ -1.38 \\ -1.51 \\ -1.89 \\ -2.96 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix} \times \begin{pmatrix} \eta \\ \tau \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \\ b_6 \\ b_7 \\ b_8 \\ b_9 \end{pmatrix} \Leftrightarrow \hat{\beta}_i = \mathbf{K}_i \times \boldsymbol{\gamma} + \mathbf{b}_i \quad i = 1, \dots, 9 \quad (5.5)$$

The Linear Mixed Effects Model

The linear mixed effects model combines both the earlier steps into a single model.

$$\mathbf{R}_i = \mathbf{X}_i \times \boldsymbol{\beta} + \mathbf{Z}_i \times \mathbf{b}_i + \boldsymbol{\epsilon}_i \quad (5.7)$$

models **fixed** component
(i.e. the relationship of
NAP to each site/all
beaches)

models **random** component
(i.e. beach is allowed to have a
different NAP-Richness
relationship)

Assumptions:

the random effects \mathbf{b}_i are normally distributed

the errors $\boldsymbol{\epsilon}_i$ (containing the five errors ϵ_{i1} to ϵ_{i5}) are normally distributed with covariance matrix $\boldsymbol{\Sigma}_i$.

$$\mathbf{b}_i \sim \mathbf{N}(\mathbf{0}, \mathbf{D})$$

$$\boldsymbol{\epsilon}_i \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$$

$$\mathbf{b}_1, \dots, \mathbf{b}_N, \boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_N \text{ independent}$$

The Random Intercept Model

$$R_{ij} = \alpha + \beta_1 \times Beach_i + \beta_2 \times NAP_{ij} + \varepsilon_{ij} \quad (5.9)$$

but high cost of degrees of freedom!

$$\begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ R_{i4} \\ R_{i5} \end{pmatrix} = \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i2} \\ 1 & NAP_{i3} \\ 1 & NAP_{i4} \\ 1 & NAP_{i5} \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \times b_i + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix} \Leftrightarrow \mathbf{R}_i = \mathbf{X}_i \times \boldsymbol{\beta} + \mathbf{Z}_i \times \mathbf{b}_i \quad (5.10)$$

this model allows your intercept to vary by beach

The Random Slope/Intercept Model

Suppose that the relationship between species richness and NAP is different on each beach....

Therefore, need to include a NAP–Beach interaction term to the model. Such a model is specified by (ANCOVA):

$$R_i = \text{factor(Beach)} + \text{NAP} \times \text{factor(Beach)}.$$

Because beach has nine levels and one level is used as the baseline, the number of parameters used by this model is excessively high, at 17. And we are not even interested in beach effects...too many df!

To estimate model degrees of freedom more efficiently, we can apply the mixed effects model with a random intercept (as before) *and* a random slope.

The Random Slope/Intercept Model

The random effects part now has three standard errors and one correlation term. The model that we are fitting is of the form

$$\begin{pmatrix} R_{i1} \\ R_{i2} \\ R_{i3} \\ R_{i4} \\ R_{i5} \end{pmatrix} = \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i2} \\ 1 & NAP_{i3} \\ 1 & NAP_{i4} \\ 1 & NAP_{i5} \end{pmatrix} \times \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} 1 & NAP_{i1} \\ 1 & NAP_{i2} \\ 1 & NAP_{i3} \\ 1 & NAP_{i4} \\ 1 & NAP_{i5} \end{pmatrix} \times \begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \varepsilon_{i3} \\ \varepsilon_{i4} \\ \varepsilon_{i5} \end{pmatrix} \quad (5.11)$$

The only difference with this model, compared to the one in Equation (5.10), is the modification of the matrix \mathbf{Z}_i ; NAP values for beach i have be included. As a result, \mathbf{b}_i is now of dimension 2×1 , and its distribution is given by

$$\begin{pmatrix} b_{i1} \\ b_{i2} \end{pmatrix} \sim N(\mathbf{0}, \mathbf{D}) \quad \text{where} \quad \mathbf{D} = \begin{pmatrix} d_{11}^2 & d_{12} \\ d_{12} & d_{22}^2 \end{pmatrix} \quad (5.12)$$