What is a hierarchical model?

HIERARCHICAL AND MIXED EFFECTS MODELS IN R

Richard Erickson
Data Scientist

datacamp

Course overview

- Components for mixed-effect models
- Applying and interpreting linear mixed-effect models
- Generalized linear mixed-effect models
- Repeated measure models

Why do we use a hierarchical model?

- Data nested within itself
- Pool information across small sample sizes
- Repeated observations across groups or individuals

Other names for hierarchical models

- Hierarchical models: Nested models, Multi-level models
- Regression framework
  - "Pool" information
  - "Random-effect" versus a "fixed-effect"
  - "Mixed-effect" (linear mixed-effect model; LMM)
  - Linear mixed-effect regression (lmer)
- Repeated sampling: "Repeated-measures", "Paired-tests"

School test scores

Meta-data:

- Gain in math scores for individual students from kindergarten to 1st grade
- Part of a national-level assessment in US
- Subset of data from West, Welch, and Galecki

Student-level variables:

- Student ID: `childid`
- Math test-score gain: `mathgain`
- Math kindergarten score: `mathdind`
- Student's sex: `sex`
- Student's minority status: `minority`

School test scores

Classroom-level variables:

- Classroom id: `classid`
- Teacher's math training: `mathprep`
- Teacher's math test knowledge test score: `mathknow`
- Teacher's years teaching: `yearstea`

School-level variables:

- School ID: `schoolid`
- School's household poverty level: `housepov`
- School's socioeconomic status: `ses`

Let's practice!

HIERARCHICAL AND MIXED EFFECTS MODELS IN R

Parts of a regression

HIERARCHICAL AND MIXED EFFECTS MODELS IN R

Richard Erickson
Data Scientist

datacamp

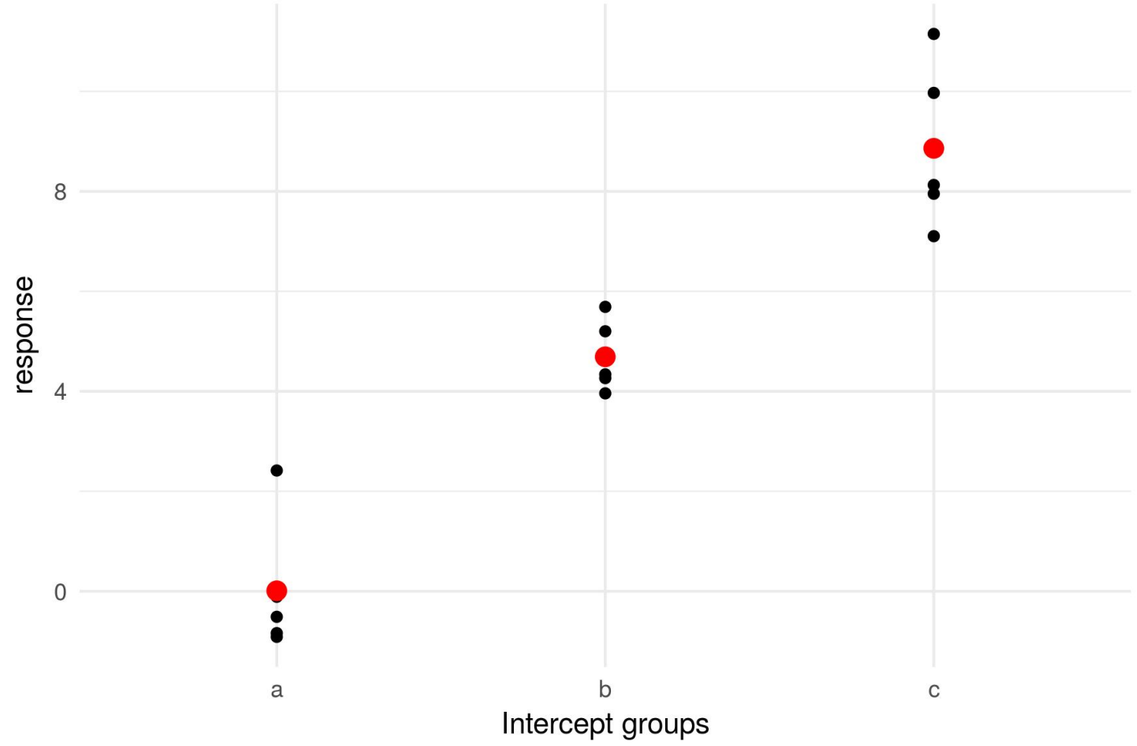
An intercept

$$y = \beta + \epsilon$$

Multiple intercepts

$$y = \beta_0 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \epsilon$$



Linear models in R

```
lm(formula, data)
```

```
lm(y ~ x, data = myData)
```

```
anova(lm(y ~ x, data = myData))
```

A simple linear regression with slopes

$$y \sim \beta_0 + \beta_1 x + \epsilon$$

Multiple regression

$$y \sim \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \epsilon$$

Multiple regression caveats

- Independence of predictor variables
- "corrected for..."
- Simpson's paradox
- Only linear
- Interactions may be important

Multiple regression in R tips

- `lm(y ~ x - 1)` estimates an intercept for each x
- Numeric versus factors
- Scaling parameters and slopes
- `lm(y ~ x1 + x2 + x1:x2)` can be written as `lm(y ~ x1 * x2)`

Refresher of running and plotting a linear regression in R

```
reg_model <- lm(response ~ predictor, data = reg_demo)
summary(reg_model)
reg_model
reg_coef_plot <- tidy(reg_model)
ggplot(reg_model, aes(x = predictor, y = response)) +
    geom_point() +
    theme_minimal()  +
    geom_abline(intercept = reg_model$estimate[1],
                slope = reg_model$estimate[2])
```

Let's practice!

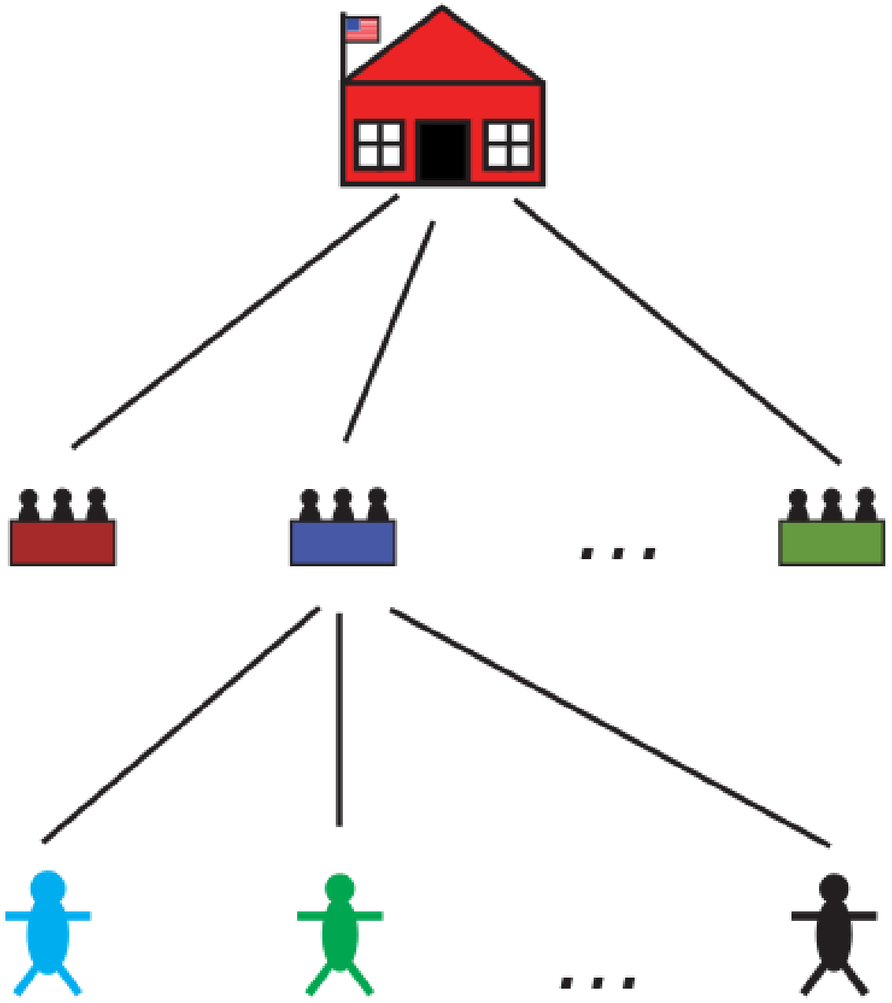
HIERARCHICAL AND MIXED EFFECTS MODELS IN R

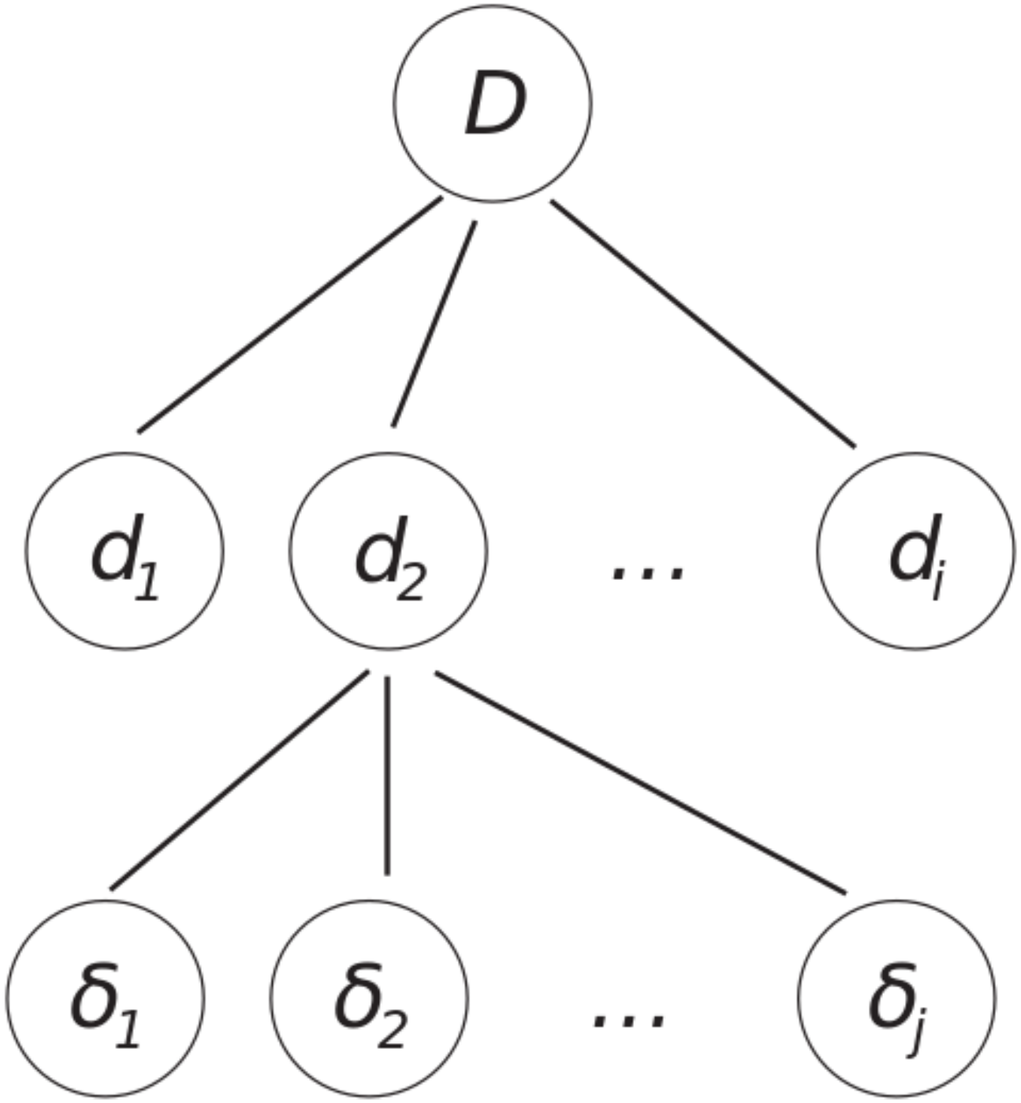
Random-effects in regressions

HIERARCHICAL AND MIXED EFFECTS MODELS IN R

Richard Erickson
Data Scientist

datacamp




D
d_1
d_2
...
d_i
δ_1
δ_2
...
δ_j

Algebraic representation

$$y \sim \beta_i x + \epsilon$$

$$\beta_i \sim \text{ Normal}(\mu, \sigma)$$

R syntax

```
library(lme4)
lmer(y ~ x + (1 | random_group), data = my_data)
lmer(y ~ x + (random_slope | random_goup),
     data = my_data)
```

Random-effect models with school data

1. Does the sex of a student impact their knowledge gain?
2. Does the teacher's training impact the gain and does the teacher's math knowledge impact the gain?

Let's practice!

HIERARCHICAL AND MIXED EFFECTS MODELS IN R