

## CHAPTER: MULTILEVEL MODELS

We have so far considered dependencies within data, when given a single set of random effects units were viewed as independent — in these situations we have a single layer of clustering. In many situations there are multiple levels of clustering, and in this case **hierarchical** or **multilevel** models can be specified to acknowledge that units within the same clusters tend to produce more similar responses.

Multilevel models are widely used in the fields of education, sociology, and demography, and are receiving increasing attention in public health and epidemiology. Multilevel models allow dependencies within data due to different levels of clustering to be modeled, but also allow the effects of variables defined at different levels to be simultaneously estimated.

For example, when between-group differences in an outcome (for example, disease rates) are attributable to differences in group composition (that is, in the characteristics of the individuals of which the groups are comprised) they are said to result from *compositional effects*. On the other hand, when group differences are attributable to the effects of group level variables or properties, they are said to result from *contextual effects*. Contextual effects often refer to the effect of a derived group-level variable (for example, mean income), but may also refer to neighborhood effects.

363

### Nested and Crossed Effects

A crucial distinction is between **nested** and **cross-classified** observations. In the simplest case we have an array of observations which we denote by  $Y_{ij}$  in which the labelling of the repeat observations within each  $i$  (indexed by  $j$ ) is essentially arbitrary, so that  $j = 1$  at  $i = 1$  has no meaningful connection with  $j = 1$  at  $i = 2$ . We say that the second suffix is **nested** within the first. For example, in a study of school performance, children are nested within schools within a school district.

By contrast,  $Y_{ij}$  can be thought of as a row by column  $A \times B$  array in which the column labelling retains the same meaning for each row  $i$  — we say that rows are **crossed** with columns. For example, we may have machines and workers crossed with machines.

364

### Example: Television School and Family Smoking Prevention and Cessation Project (TVSFP)

Reference: Flay et al. (1995), reanalyzed by Fitzmaurice, Laird and Ware (2004).

The TVSFP was a study designed to determine the efficacy of a school-based smoking prevention curriculum in conjunction with a television-based prevention program, in terms of preventing smoking onset and increasing smoking cessation.

The study used a  $2 \times 2$  factorial design, with four intervention conditions determined by the cross-classification of a school-based social-resistance curriculum (CC: coded 1 = yes, 0 = no) with a television-based prevention program (TV: coded 1 = yes, 0 = no). Randomization to one of the four intervention conditions was at the **school** level, while much of the intervention was delivered at the classroom level. The original study involved 6695 students in 47 schools in Southern California. This dataset consists of a subset of 1600 seventh-grade students from 135 classes in 28 schools in Los Angeles. The response variable, a tobacco and health knowledge scale (THKS), was administered before and after randomization of schools to one of the four intervention conditions. The scale assessed a student's knowledge of tobacco and health — high numbers are better.

365

```
TVdat <- read.table("TV.dat",col.names=c("schoolID","classID","CC","TV","Pre.THKS","Y"))
> TVdat[1:5,]
  schoolID classID CC TV Pre.THKS Y
1      403  403101  1  0         2 3
2      403  403101  1  0         4 4
3      403  403101  1  0         4 3
4      403  403101  1  0         3 4
5      403  403101  1  0         3 4
> TVdat[1595:1600,]
  schoolID classID CC TV Pre.THKS Y
1595     515  515113  0  0         2 0
1596     515  515113  0  0         2 1
1597     515  515113  0  0         1 1
1598     515  515113  0  0         2 1
1599     515  515113  0  0         3 2
1600     515  515113  0  0         3 3
> dim(TVdat)
[1] 1600    6
```

366

## Models for the TVSFP Data

We consider a linear model for the post-intervention THKS score, with the pre-intervention THKS score as a covariate. We include main effects for CC and TV, along with the interaction.

Here we have three level data: individuals (level 1), classrooms (level 2) and schools (level 3).

School and classroom effects were incorporated via the introduction of random effects at levels 3 and 2 respectively. We let  $Y_{ijk}$  be the post-intervention score for the  $k$ -th student within the  $j$ -th classroom in school  $i$ ;  $i = 1, \dots, 28$ ,  $j = 1, \dots, 135$ ,  $k = 1, \dots, 1600$ .

Hence we have the model:

$$Y_{ijk} = \beta_0 + \beta_1 \text{Pre.THKS}_{ijk} + \beta_2 \text{CC}_i + \beta_3 \text{TV}_i + \beta_4 \text{CC}_i \times \text{TV}_i + b_i + b_{ij} + \epsilon_{ijk}$$

where  $b_i \sim_{iid} N(0, \sigma_3^2)$ ,  $b_{ij} \sim_{iid} N(0, \sigma_2^2)$  and  $\epsilon_{ijk} \sim_{iid} N(0, \sigma_1^2)$  with  $b_i, b_{ij}, \epsilon_{ijk}$  independent.

367

## Marginal Moments

Marginal mean:

$$E[Y_{ijk}] = \beta_0 + \beta_1 \text{Pre.THKS}_{ijk} + \beta_2 \text{CC}_i + \beta_3 \text{TV}_i + \beta_4 \text{CC} \times \text{TV}_i$$

Marginal correlation between scores on classmates is:

$$\text{corr}(Y_{ijk}, Y_{ijk'}) = \frac{\sigma_2^2 + \sigma_3^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}$$

Marginal correlation between scores on children from different class rooms in the same school is:

$$\text{corr}(Y_{ijk}, Y_{ij'k'}) = \frac{\sigma_3^2}{\sigma_1^2 + \sigma_2^2 + \sigma_3^2}$$

We now fit this model — note the use of the / symbol.

368

```

> TVmod1 <- lme(Y~Pre.THKS+CC+TV+CC:TV,data=TVdat,random=~1|schoolID/classID)
> summary(TVmod1)
Linear mixed-effects model fit by REML
Random effects:
  Formula: ~1 | schoolID
            (Intercept)
StdDev:    0.1965707
  Formula: ~1 | classID %in% schoolID
            (Intercept) Residual
StdDev:    0.2542862 1.265818
Fixed effects: Y ~ Pre.THKS + CC + TV + CC:TV
              Value Std.Error  DF  t-value p-value
(Intercept)  1.7019852 0.12543005 1464 13.569198  0.0000
Pre.THKS     0.3053628 0.02589132 1464 11.794021  0.0000
CC           0.6413260 0.16094730   24  3.984696  0.0005
TV           0.1820802 0.15724055   24  1.157972  0.2583
CC:TV       -0.3309400 0.22458558   24 -1.473559  0.1536
Number of Observations: 1600
Number of Groups:
      schoolID classID %in% schoolID
           28           135

```

369

Let's drop the interaction.

```

> TVmod2 <- lme(Y~Pre.THKS+CC+TV,data=TVdat,random=~1|schoolID/classID)
> summary(TVmod2)
andom effects:
  Formula: ~1 | schoolID
            (Intercept)
StdDev:    0.1981275
  Formula: ~1 | classID %in% schoolID
            (Intercept) Residual
StdDev:    0.261783 1.265338
              Value Std.Error  DF  t-value p-value
(Intercept)  1.7849270 0.11294585 1464 15.803387  0.0000
Pre.THKS     0.3052393 0.02589749 1464 11.786441  0.0000
CC           0.4714724 0.11330446   25  4.161110  0.0003
TV           0.0195612 0.11330016   25  0.172649  0.8643

```

370

Now let's drop TV from the model.

```
> TVmod3 <- lme(Y~Pre.THKS+CC,data=TVdat,random=~1|schoolID/classID)
> summary(TVmod3)
Random effects:
  Formula: ~1 | schoolID
            (Intercept)
StdDev:    0.1891069
  Formula: ~1 | classID %in% schoolID
            (Intercept) Residual
StdDev:    0.2616994 1.265362
Fixed effects: Y ~ Pre.THKS + CC
              Value Std.Error DF t-value p-value
(Intercept) 1.7934267 0.09495984 1464 18.886160 0e+00
Pre.THKS     0.3055676 0.02588490 1464 11.804859 0e+00
CC           0.4723229 0.11088966 26 4.259395 2e-04
> TVmod4 <- lm(Y~Pre.THKS+CC+TV+CC:TV,data=TVdat)
> summary(TVmod4)
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.66126    0.08436  19.693 < 2e-16 ***
Pre.THKS     0.32518    0.02585  12.578 < 2e-16 ***
CC           0.64055    0.09210   6.955 5.14e-12 ***
TV           0.19871    0.08996   2.209 0.0273 *
CC:TV       -0.32162    0.13025  -2.469 0.0136 *
```

371

The estimated marginal correlation between scores on classmates is:

$$\text{corr}(Y_{ijk}, Y_{ijk'}) = \frac{\hat{\sigma}_2^2 + \hat{\sigma}_3^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^2} = \frac{0.262^2 + 0.189^2}{1.265^2 + 0.262^2 + 0.189^2} = 0.061$$

Estimated marginal correlation between scores on children from different classes in the same school is:

$$\text{corr}(Y_{ijk}, Y_{ij'k'}) = \frac{\hat{\sigma}_3^2}{\hat{\sigma}_1^2 + \hat{\sigma}_2^2 + \hat{\sigma}_3^2} = \frac{0.189^2}{1.265^2 + 0.262^2 + 0.189^2} = 0.021$$

These are small — but it makes a huge difference to conclusions to ignore them as we see from model TVmod4 in which the interaction and both main effects are significant.

### Example: Childhood outcomes in Guatemala

In the final you will analyze data from a study in which the binary outcome was a full set of immunizations, given at least some immunization. The aim was to see how this outcome varies with individual, family and community variables.

In this study there were 2159 children aged 1–4, with multiple pregnancies per mother (1595 mothers), and across 161 communities. We might expect that outcomes from the same mother to be correlated, and outcomes within the same community to be correlated.

Covariates can be measured at various levels:

#### Individual

- Child aged  $\geq 2$  years.
- Mother aged  $\geq 25$ .
- Birth order 2–3.
- Birth order 4–6.
- Birth order  $\geq 7$ .

373

#### Family

- Indigenous, no Spanish.
- Indigenous, Spanish.
- Mother's education primary.
- Mother's education secondary or better.
- Husband's education primary.
- Husband's education secondary or better.
- Husband's education missing.
- Mother ever worked.

#### Community

- Rural.
- Proportion indigenous, 1981.

374

Let  $p_{ijk}$  be the probability of a full set of immunizations, given at least some on the  $k$ -th child in the  $j$ -th family in the  $i$ -th community.

We may then consider models of the form

$$\log \left( \frac{p_{ijk}}{1 - p_{ijk}} \right) = \beta_0 + \beta_1 \mathbf{x}_{ijk} + \beta_2 \mathbf{x}_{ij} + \beta_3 \mathbf{x}_i + b_{ij} + b_i$$

where  $\mathbf{x}_{ijk}$ ,  $\mathbf{x}_{ij}$  and  $\mathbf{x}_i$  represent vectors of covariates at the individual, family and community levels, respectively, with corresponding fixed effects  $\beta_1$ ,  $\beta_2$  and  $\beta_3$ ;  $b_{ij} \sim_{iid} N(0, \sigma_2^2)$  and  $b_i \sim_{iid} N(0, \sigma_3^2)$  are random effects representing unmeasured variables at the family and community levels, with  $b_{ij}$  and  $b_i$  independent.

375

### Split Plot Designs

In a split plot design there are two factors, one that is easy to change (call this factor B), and one that is more difficult (call this factor A).

An example is in agriculture where factor A represents irrigation levels which can only be applied to large plots of land, and B represents different crop varieties.

Levels of A are randomly assigned to whole plots while levels of B are randomly assigned to **split plots** (subplots) within each plot.

376

**Example: Oats**

Yates (1935) introduced split-plot data in a  $3 \times 4$  full factorial design with 3 varieties of oats and four concentrations on nitrogen.

The experimental units were arranged into 6 blocks, each with 3 whole-plots subdivided 4 subplots. The varieties of oats were assigned randomly to the whole-plots and the concentrations of nitrogen to the subplots. Hence there are  $6 \times 3 \times 4$  observations in total.

Pinheiro and Bates (2000) analyze these data — they are available as data `Oats`. The subplot `yield` is in bushels/acre, `nitro` is the concentration in cwt/acre (0.0, 0.2, 0.4, 0.6), `Block` is a factor labeled I, II, III, IV, V, VI, and `Variety` is one of Golden Rain, Marvellous, Victory.

In terms of dependencies we would expect yields within the same block to be correlated, and also yields within the same subplot to be correlated. Subplots are nested within blocks.

377

```
> library(nlme)
> data(Oats)
Grouped Data: yield ~ nitro | Block
  Block  Variety nitro yield
1     I  Victory  0.0  111
2     I  Victory  0.2  130
3     I  Victory  0.4  157
4     I  Victory  0.6  174
5     I Golden Rain  0.0  117
6     I Golden Rain  0.2  114
7     I Golden Rain  0.4  161
8     I Golden Rain  0.6  141
9     I  Marvellous  0.0  105
....
62    VI  Victory  0.2   74
63    VI  Victory  0.4  118
64    VI  Victory  0.6  113
65    VI Golden Rain  0.0   89
66    VI Golden Rain  0.2   82
67    VI Golden Rain  0.4   86
68    VI Golden Rain  0.6  104
69    VI  Marvellous  0.0   97
70    VI  Marvellous  0.2   99
71    VI  Marvellous  0.4  119
72    VI  Marvellous  0.6  121
```

378



### Visualizing the Split-Plot Data

Figure 38 was produced using the command:

```
plot(Oats,display="Block",inner=~Variety)
```

We see that the yields clearly increase on average for all 3 varieties, and the trend seems to be similar for all 3.

Other than that no one variety is obviously better, though there are clear differences between blocks, with block I roducing higher yields.

379

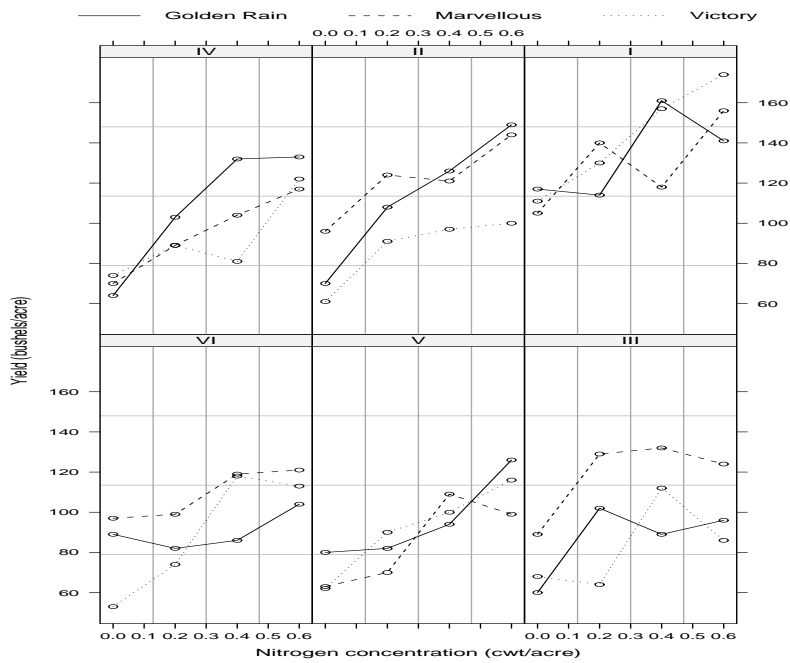


Figure 38: Oat yields by block and variety.

## Models for the Split-Plot Data

We have three levels of grouping here corresponding to block, plot and subplot — we may be tempted to associate a random effect with each but there is only one yield per subplot and so we would saturate the model with random effects.

We can use random intercepts for blocks and whole plots nested within blocks. Let  $Y_{ijk}$  be the yield on nitrogen level  $x_{ijk}$  with variety  $j$  ( $j = 1/2/3$  corresponding to Golden Rain/Marvellous/Victory) within block  $i$ ,  $i = 1, \dots, 6$  and  $V_{1ijk}$  be an indicator for variety Marvellous and  $V_{2ijk}$  be an indicator for variety Victory.

381

Consider the model

$$Y_{ijk} = \beta_0 + \beta_1 x_{ijk} + \beta_2 V_{1ijk} + \beta_3 V_{2ijk} + b_i + b_{ij} + \epsilon_{ijk}$$

with

$$\begin{aligned} b_i &\sim_{iid} N(0, \sigma_0^2) \\ b_{ij} &\sim_{iid} N(0, \sigma_1^2) \\ \epsilon_{ijk} &\sim_{iid} N(0, \sigma_\epsilon^2) \end{aligned}$$

with  $b_i, b_{ij}, \epsilon_{ijk}$  independent, for  $i = 1, \dots, 6$  blocks,  $j = 1, 2, 3$  varieties and  $k = 1, \dots, 4$ ;

- $\beta_0$  is the mean for Golden Rain (at the lowest level of nitrogen),
- $\beta_1$  is the effect of level of nitrogen,
- $\beta_2$  is the change in yield in moving from Golden Rain to Marvellous,
- $\beta_3$  is the change in yield in moving from Golden Rain to Victory.

382

```

> mod0 <- lme(yield~nitro+Variety,data=Oats,random=~1|Block/Variety)
> summary(mod0)
Linear mixed-effects model fit by REML
Data: Oats
      AIC      BIC    logLik
592.8918 608.4283 -289.4459
Random effects:
Formula: ~1 | Block
      (Intercept)
StdDev:    14.64483
Formula: ~1 | Variety %in% Block
      (Intercept) Residual
StdDev:    10.43758 12.86697
Fixed effects: yield ~ nitro + Variety
              Value Std.Error DF   t-value p-value
(Intercept)  82.40000  8.058511 53 10.225214  0.0000
nitro        73.66667  6.781487 53 10.862908  0.0000
VarietyMarvellous  5.29167  7.078907 10  0.747526  0.4720
VarietyVictory   -6.87500  7.078907 10 -0.971195  0.3544

```

We drop varieties from the model.

383

```

> mod0M <- lme(yield~nitro+Variety,data=Oats,random=~1|Block/Variety,method="ML")
> mod1M <- lme(yield~nitro,data=Oats,random=~1|Block/Variety,method="ML")
> anova(mod0M,mod1M)
      Model df      AIC      BIC    logLik  Test  L.Ratio p-value
mod0M     1  7 615.1077 631.0444 -300.5539
mod1M     2  5 614.2290 625.6123 -302.1145 1 vs 2 3.121277  0.21
> mod1 <- lme(yield~nitro,data=Oats,random=~1|Block/Variety)
> summary(mod1)
Data: Oats
      AIC      BIC    logLik
603.0418 614.2842 -296.5209
Random effects:
Formula: ~1 | Block
      (Intercept)
StdDev:    14.50604
Formula: ~1 | Variety %in% Block
      (Intercept) Residual
StdDev:    11.0047 12.86695

Fixed effects: yield ~ nitro
              Value Std.Error DF   t-value p-value
(Intercept) 81.87222  6.945302 53 11.78814    0
nitro       73.66667  6.781478 53 10.86292    0
> plot(augPred(mod1),layout=c(6,3))

```

384

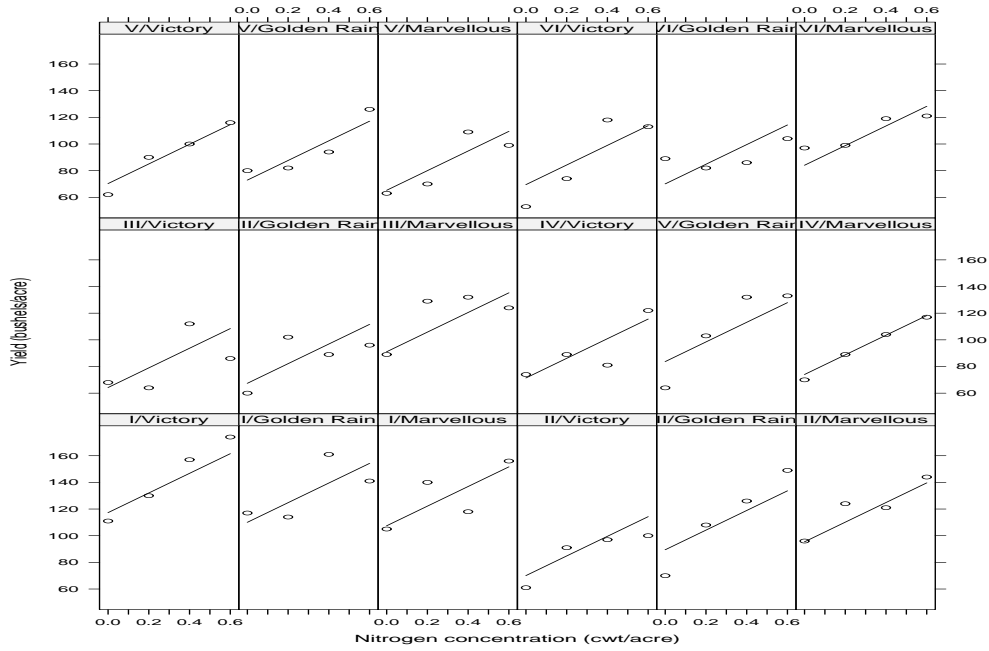


Figure 39: Fitted curves from mod1.

385

Oats: Random Effect Predictions

```
> ranef(mod1,level=1:2)
Level: Variety %in% Block      Level: Block
      (Intercept)              (Intercept)
VI/Golden Rain  -5.6498642    VI      -6.141400
VI/Marvellous   8.3242076     V       -10.382943
VI/Victory      -6.2088270    III      -6.406496
V/Golden Rain   1.4240290     IV       -4.617096
V/Marvellous    -6.2151303    II       2.606781
V/Victory       -1.1844644    I       24.941154
III/Golden Rain -8.0607855
III/Marvellous  15.6019761
III/Victory     -11.2282418
IV/Golden Rain  6.4428844
IV/Marvellous   -3.2458053
IV/Victory      -5.8542987
II/Golden Rain  4.9717858
II/Marvellous   10.9340564
II/Victory      -14.4055938
I/Golden Rain   3.2320159
I/Marvellous    0.6235225
I/Victory       10.4985332
```

386