Regression: Mixed-effects approach

A mixed model is a statistical model containing both **fixed** effects and **random** effects. These models are useful in a wide variety of disciplines in the physical, biological and social sciences. They are particularly useful in settings where repeated measurements are made on the same statistical units (longitudinal study), or where measurements are made on clusters of related statistical units. Because of their advantage in dealing with missing values, mixed effects models are often preferred over more traditional approaches.

Definition(s) of mixed-effects: Confusion!

1. Fixed effects are constant across individuals, and random effects vary. For example, in a growth study, a model with random intercepts α_i and fixed slope β corresponds to parallel lines for different individuals i, or the model $y_{it} = \alpha_i + \beta t$. Kreft and de Leeuw [(1998), page 12] thus distinguish between fixed and random coefficients.
2. Effects are fixed if they are interesting in themselves or random if there is interest in the underlying population. Searle, Casella and McCulloch [(1992), Section 1.4] explore this distinction in depth.
3. "When a sample exhausts the population, the corresponding variable is *fixed*; when the sample is a small (i.e., negligible) part of the population the corresponding variable is *random*" [Green and Tukey (1960)].
4. "If an effect is assumed to be a realized value of a random variable, it is called a random effect" [LaMotte (1983)].
5. Fixed effects are estimated using least squares (or, more generally, maximum likelihood) and random effects are estimated with shrinkage ["linear unbiased prediction" in the terminology of Robinson (1991)]. This definition is standard in the multilevel modeling literature [see, e.g., Snijders and Bosker (1999), Section 4.2] and in econometrics.

Gelman, A. (2005). Analysis of variance - why it is more important than ever? The Annals of Statistics, 33(1), 1–53.
http://doi.org/10.1214/009053604000001048

Definition(s) of mixed-effects: Confusion!

We prefer to sidestep the overloaded terms "fixed" and "random" with a cleaner distinction by simply renaming the terms in definition 1 above. We define effects (or coefficients) in a multilevel model as *constant* if they are identical for all groups in a population and *varying* if they are allowed to differ from group to group. For example, the model $y_{ij} = \alpha_j + \beta x_{ij}$ (of units i in groups j) has a constant slope and varying intercepts, and $y_{ij} = \alpha_j + \beta_j x_{ij}$ has varying slopes and intercepts. In this terminology (which we would apply at any level of the hierarchy in a multilevel model), varying effects occur in batches, whether or not the effects are interesting in themselves (definition 2), and whether or not they are a sample from a larger set (definition 3). Definitions 4 and 5 do not arise for us since we estimate all batches of effects hierarchically, with the variance components σ_m estimated from data.

Gelman, A. (2005). Analysis of variance - why it is more important than ever? The Annals of Statistics, 33(1), 1–53.
http://doi.org/10.1214/009053604000001048

Mixed-effects approach: Notation

units i in groups j

$$y_{ij} = \alpha_j + \beta x_{ij}$$

varying intercepts constant slope

$$y_{ij} = \alpha_j + \beta_j x_{ij}$$

varying intercepts varying slopes

Gelman, A. (2005). Analysis of variance - why it is more important than ever? The Annals of Statistics, 33(1), 1–53.
http://doi.org/10.1214/009053604000001048

reasons for using mixed-effects: Nested structures

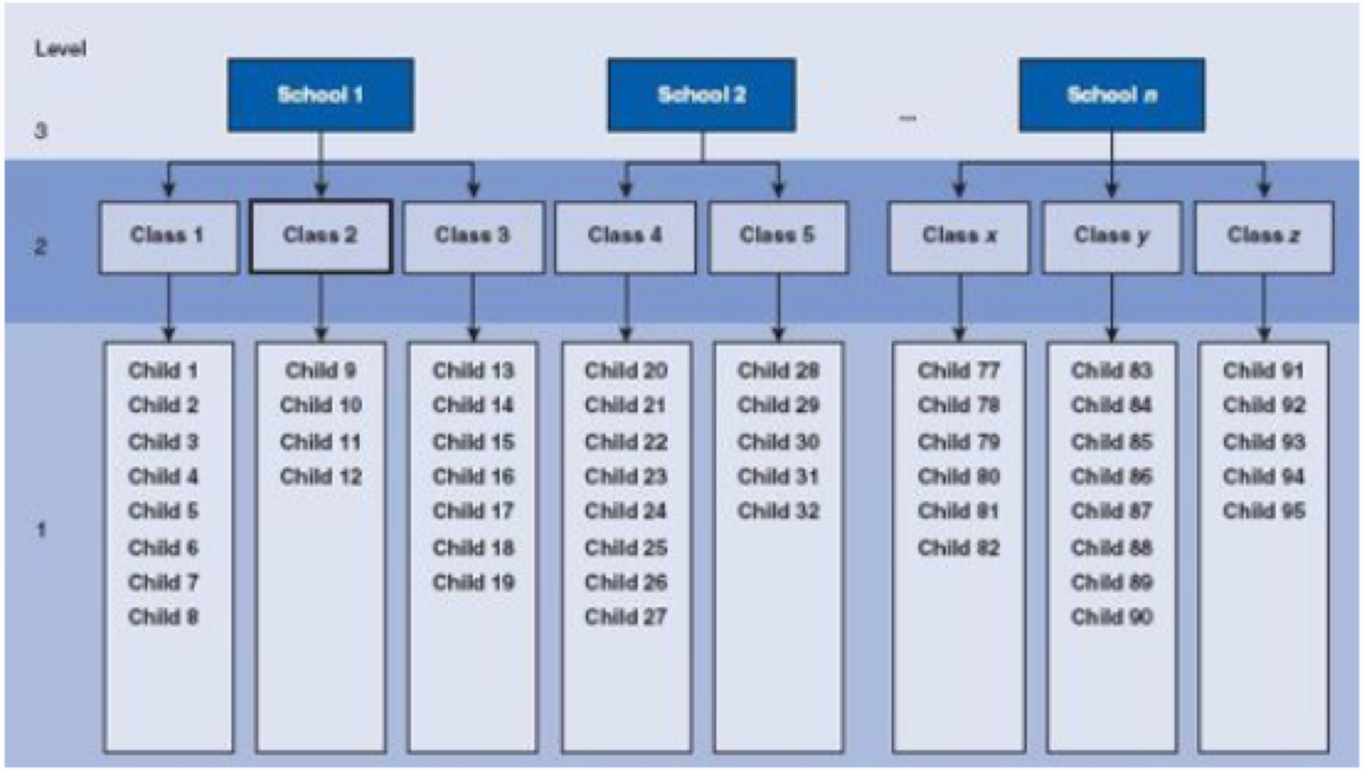
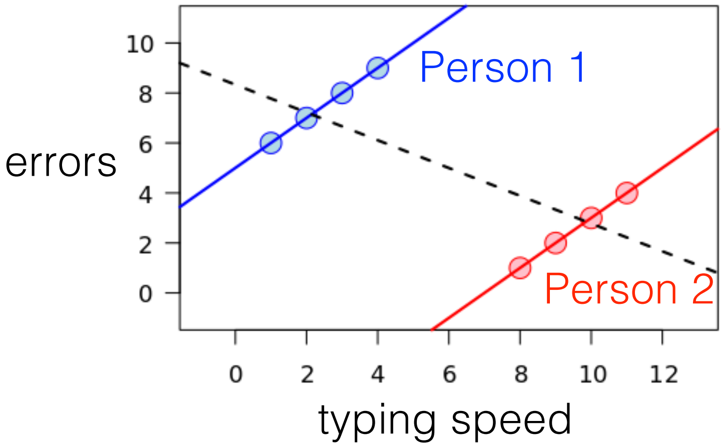


FIGURE 19.3 An example of a three-level hierarchical data structure

Field, A., Miles, J., & Field, Z (2012). *Discovering statistics using R*. SAGE Publications.

reasons for using mixed-effects: Simpsons' paradox

Paradox in which a trend that appears in groups of data disappears when these groups are combined and the reverse trend appears for the aggregate data.



speed-accuracy trade-off

people who are faster are also more accurate (between)

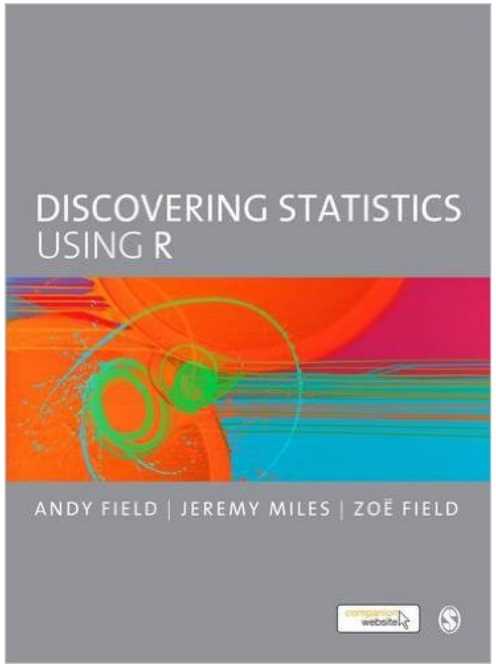
when people are faster they become less accurate (within)

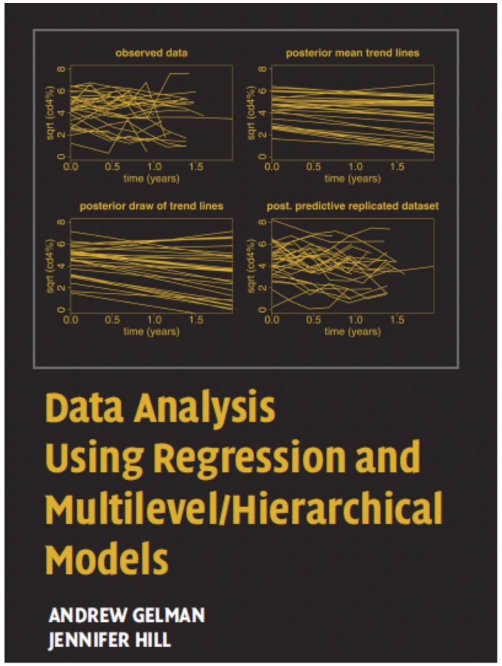
exercise and fatigue

people who exercise more are less fatigued (between)

when people exercise more they are more fatigued (within)

suggested readings





Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. http://doi.org/10.1016/j.jml.2007.12.005

Judd, C. M., Westfall, J., & Kenny, D. A. (2012). Treating stimuli as a random factor in social psychology: A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology, 103*(1), 54–69. http://doi.org/10.1037/a0028347

Fitting Linear Mixed-Effects Models Using **lme4**

```
model <- lmer(Y ~ x + (x | g), data)
```

<table>
<tr><th>Formula</th><th>Alternative</th><th>Meaning</th></tr>
<tr><td><code>(1 | g)</code></td><td><code>1 + (1 | g)</code></td><td>Random intercept with fixed mean.</td></tr>
<tr><td><code>0 + offset(o) + (1 | g)</code></td><td><code>-1 + offset(o) + (1 | g)</code></td><td>Random intercept with a priori means.</td></tr>
<tr><td><code>(1 | g1/g2)</code></td><td><code>(1 | g1)+(1 | g1:g2)</code></td><td>Intercept varying among <code>g1</code> and <code>g2</code> within <code>g1</code>.</td></tr>
<tr><td><code>(1 | g1) + (1 | g2)</code></td><td><code>1 + (1 | g1) + (1 | g2).</code></td><td>Intercept varying among <code>g1</code> and <code>g2</code>.</td></tr>
<tr><td><code>x + (x | g)</code></td><td><code>1 + x + (1 + x | g)</code></td><td>Correlated random intercept and slope.</td></tr>
<tr><td><code>x + (x || g)</code></td><td><code>1 + x + (1 | g) + (0 + x | g)</code></td><td>Uncorrelated random intercept and slope.</td></tr>
</table>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software, 67(1), 1–51. http://doi.org/10.18637/jss.v067.i01