Multilevel Models (MLM)

- "Multilevel modeling is a generalization of generalized linear modeling" (Gelman, 2005)
- Multilevel model = Hierarchical model = Mixed Effect model = …
- "Level" in multilevel refers to the hierarchy of parameters
  - Usually refers to a hierarchy of groupings of the individual entities

When do we use, and Why do we need it?

- Example: We want to study alcohol abuse among young people (Dominici, 2005)
  - A person is a member of a family, and a resident of a state
  - Level 1 (person): person's ability to metabolize alcohol
  - Level 2 (family): alcohol abuse in the family
  - Level 3 (state): state laws
- Multilevel modeling is useful when
  - Not enough data for lowest level models while higher level models are too coarse
  - or desire to get similar results for individuals within a group
- MLM models entities at the lowest level, but "borrows strength from" higher levels

When does it make a difference?

Classical regression vs. Multilevel modeling

- When there is very little group-level variation, multilevel modeling reduces to classical regression with no group indicators
- When group-level coefficients vary greatly, multilevel modeling reduces to classical regression with group indicators (group dummy codes)

So advantageous “in-between”, or when domain knowledge or business process dictates that certain parameters be softly “tied together”

Alcohol Use among Adolescents

- Three years of longitudinal data
- 82 adolescents, beginning at age 14
- Covariates:
  - COA: an indicator variable whether the adolescent is a child of alcoholic parent
  - PEER: 8-point scale that shows the proportion of their friends who drink alcohol
  - time: 0, 1, 2 (three years)
- 2 Levels: individuals ($i=1,\ldots 82$), and whole group, indexed by $i=0$

Proposed Multilevel Model

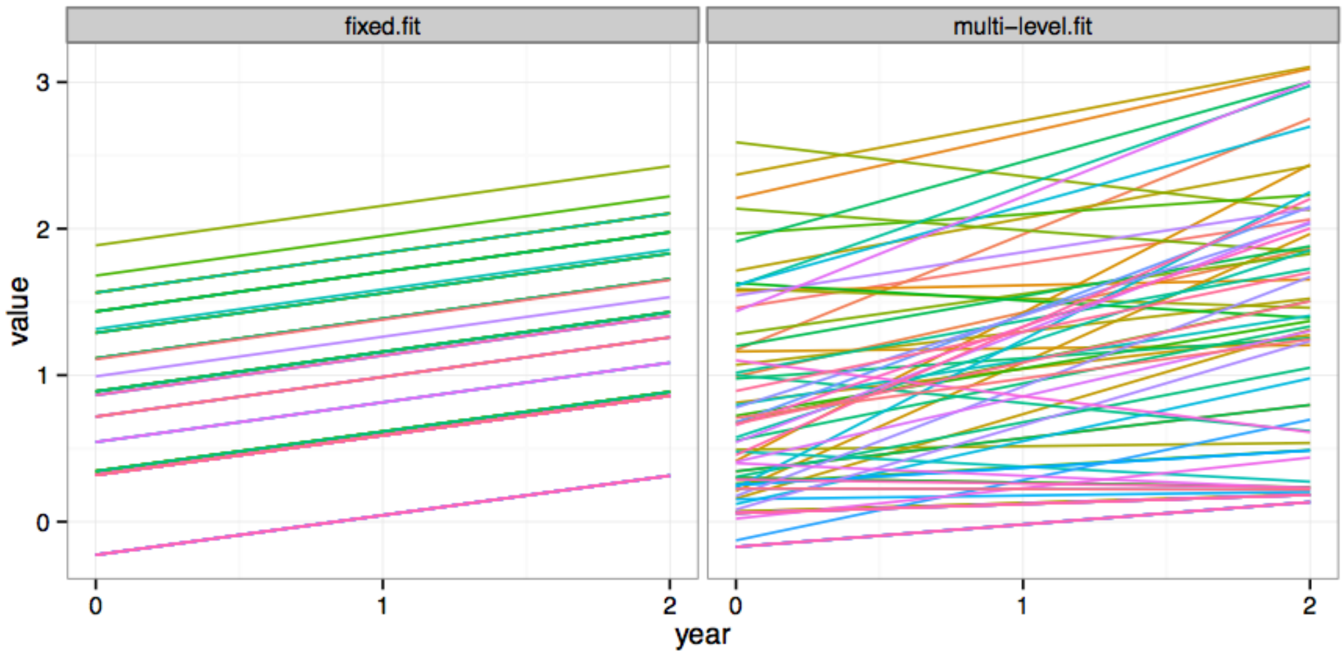
$$
\begin{aligned}
y_{it} &= \beta_{0i} + \beta_p p_{it} + \beta_c c_{it} + \beta_{1i} t + \epsilon_{it} \\
\beta_{0i} &= \beta_{00} + b_{0i} \\
\beta_{1i} &= \beta_{10} + b_{1i} \\
(b_{0i}, b_{1i} &\text{ are noise terms (called random effects; to contrast with} \\
&\beta\text{s, which are "fixed effects")} \\
y_{it} : &\text{ Alcohol use} \\
c_{it} : &\text{ COA} \\
p_{it} : &\text{ PEER} \\
t : &\text{ time}
\end{aligned}
$$

What is the corresponding MLR?

In R, using package(nlme)

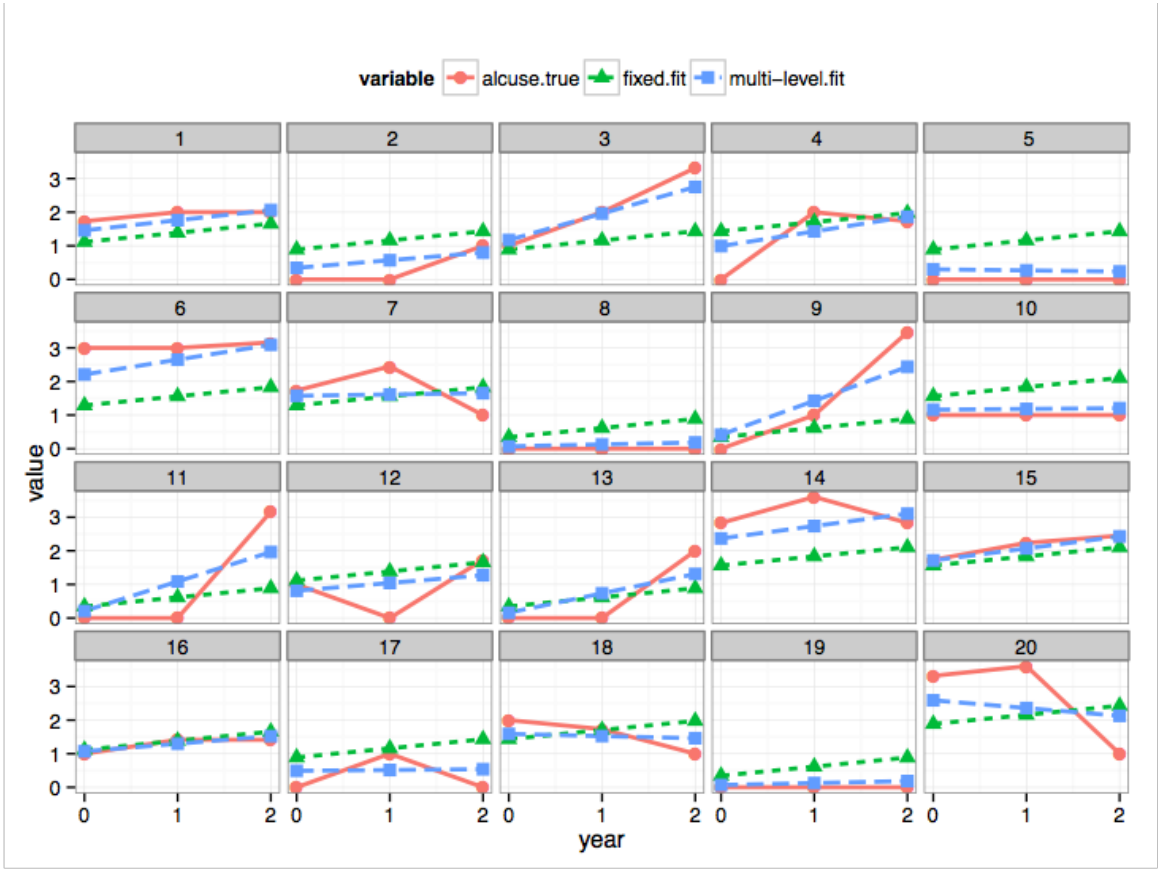
```
> model.e <- lme(alcuse ~ coa+peer+time , data=df, random= ~ time | id, method="REML")
> summary(model.e)
Linear mixed-effects model fit by REML
 Data: df
       AIC      BIC    logLik
  619.7211 647.6326 -301.8606

Random effects:
 Formula: ~time | id
 Structure: General positive-definite, Log-Cholesky parametrization
            StdDev    Corr
(Intercept) 0.5119870 (Intr)
time        0.3938927 -0.075
Residual    0.5807686

Fixed effects: alcuse ~ coa + peer + time
                 Value  Std.Error  DF   t-value p-value
(Intercept) -0.2263521 0.14036356 163 -1.612613  0.1088
coa          0.5711970 0.14897721  79  3.834123  0.0003
peer         0.6092227 0.10226024  79  5.957572  0.0000
time         0.2706514 0.06283908 163  4.307056  0.0000
 Correlation:
     (Intr) coa    peer
coa  -0.359
peer -0.664 -0.162
time -0.254  0.000  0.000

Standardized Within-Group Residuals:
       Min         Q1        Med         Q3        Max
-2.5999504 -0.3984809 -0.1047824  0.3732800  2.3604876

Number of Observations: 246
Number of Groups: 82
```

Alcohol Use Example (continued)



[Alcohol use among adolescents] Original data, and fitted curves from multilevel and classical regression. The multilevel model captures the individual-differences on the changes of alcohol use.

Alcohol Use Example (continued)



[Alcohol use among adolescents] Original data, and fitted curves from multilevel and classical regression. The multilevel model captures the individual-differences on the changes of alcohol use.

References for MLM

- (nlme in R) http://cran.r-project.org/web/packages/nlme/nlme.pdf
- (lme4 in R) http://cran.r-project.org/web/packages/lme4/lme4.pdf
- (Dominici, 2005) http://www.biostat.jhsph.edu/bstcourse/bio607/
- (Bates and Pinheiro) http://www.stat.bell-labs.com/NLME/CompMulti.pdf
- (Papers on multilevel modeling) http://www.ats.ucla.edu/stat/papers/mlmpapers.htm
- (Gelman, 2005) Gelman, Andrew. "Multilevel (hierarchical) modeling: what it can and cannot do." *Technometrics* 48.3 (2006). http://www.cs.berkeley.edu/~russell/classes/cs294/f05/papers/gelman-2005.pdf

- PYTHON implementation: part of PyMC3
  - https://pymc-devs.github.io/pymc3/GLM-hierarchical/
  - While My MCMC Gently Samples

Extras

Statistical Details*

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\epsilon}_i$$
$$\mathbf{b}_i \sim \mathrm{N}(0, \sigma^2\mathbf{D})$$
$$\boldsymbol{\epsilon}_i \sim \mathrm{N}(0, \sigma^2\mathbf{I})$$

$$\begin{aligned}
&L(\boldsymbol{\beta}, \mathbf{D}, \sigma^2 \mid \mathbf{Y}) \\
&= \prod p(\mathbf{y}_i \mid \boldsymbol{\beta}, \mathbf{D}, \sigma^2) \\
&= \prod \int p(\mathbf{y}_i \mid \mathbf{b}_i, \boldsymbol{\beta}, \sigma^2) p(\mathbf{b}_i \mid \mathbf{D}, \sigma^2) d\mathbf{b}_i \\
&= \prod \frac{1}{\sqrt{(2\pi\sigma^2)^{n_i}|\mathbf{D}|}} \int \frac{\exp \frac{-1}{2\sigma^2}(\|\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta} - \mathbf{Z}_i\mathbf{b}_i\| + \mathbf{b}_i^\top \mathbf{D}^{-1}\mathbf{b}_i)}{(2\pi\sigma^2)^{q/2}} d\mathbf{b}_i
\end{aligned}$$

Computational Details*

Takeaway: There are several ways of estimating the parameters (given as an option in the R code), all are somewhat involved compared to MLR.

[*While My MCMCGently Samples*]

- Maximum Likelihood Estimation
  - with respect to all the parameters
  - EM iterations
- Restricted Maximum Likelihood Estimation
  - focusing only on D (random-effect covariance) (Harville, 1976)
  - EM iterations
- Bayesian Posterior Estimation
  - Gibbs sampling using R and Bugs (Gelman and Hill, 2007)